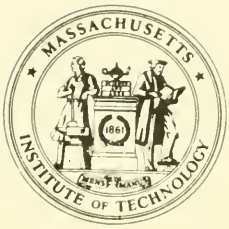


LIBRARY
OF THE
MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

79-74

MASS. INST. TECH.
NOV 20 '74
LIBRARY

WORKING PAPER
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

UNCERTAINTY IN THE OIL AND GAS EXPLORATORY PROCESS: A FRAMEWORK FOR ANALYSIS AND ECONOMETRIC ESTIMATION

by

Krishna Challa* & Marti G. Subrahmanyam*

719-74 July 1974

MASSACHUSETTS
INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139

MASS. INST. TECH.
NOV 20 '74
DEWEY LIBRARY

UNCERTAINTY IN THE OIL AND GAS EXPLORATORY PROCESS: A FRAMEWORK FOR ANALYSIS AND ECONOMETRIC ESTIMATION

by

Krishna Challa* & Marti G. Subrahmanyam*

719-74 July 1974

* Sloan School of Management, Massachusetts Institute of Tehnology

ACKNOWLEDGEMENT

We would like to express our most sincere gratitude to the inspiration and guidance we got during the entire course of this work from Professors Paul MacAvoy and Robert Pindyck. We also owe a special thanks to Kevin Lloyd and Phil Sussman for their help.

ABSTRACT

This paper attempts to provide a unified framework that takes into account various aspects of the oil and natural gas exploratory process such as the depletion of the resource base with time, uncertainty in drilling, changing economic incentives and the joint product aspects of the oil and gas exploration. The equilibrium in the market for oil and natural gas prospects is first described and the individual explorer's decision process is analyzed in the light of the the characteristics of the equilibrium. This analysis leads us to anticipations regarding the nature of economic relationships satisfied by the relevant variables. The goal here has been to derive relationships that follow naturally from a description of the choice behavior of an individual explorer. An attempt has also been made to avoid econometric biases that may have crept into some of the earlier estimations. All the implied economic relationships have been estimated and appear in the last section of the paper.

Introduction

The economics of the natural gas industry in the United States has been the subject of much attention in recent years. Producer prices of natural gas have been under area rate regulation by the Federal Power Commission following the Supreme Court decision in the Phillips case in 1954.[1] The responsiveness of the supply of natural gas to economic incentives has been the specific issue of much of the debate in recent years, particularly in view of the increasing gap between the supply and demand for natural gas. Following a pioneering study by Fisher (5), the exploratory process in the petroleum industry has been analyzed by several authors including MacAvoy (13), Erickson and Spann (3, 4, 19) Khazzoom (9,10), and MacAvoy and Pindyck (14). The key point brought out in several of these discussions is the distinction between the response of wildcat drilling to economic incentives and that of the supply of new reserves, the latter being considerably less than the former.

The exploratory process is triggered off by increased geophysical activity in response to positive economic incentives, followed by exploratory drilling of various types.[2] Depending on the intent of exploratory drilling (oil, gas, or hydrocarbon), the 'directionality,'[3] and type of drilling undertaken, the outcome of the drilling process

---

[1] Phillips Petroleum Co. v. Wisconsin, 347 U.S. 672 (1954)

[2] Exploratory drilling activity has been classified by the American Association of Petroleum Geologists (AAPG) into several categories--new field wildcats, new pool wildcats, deeper pool tests, shallower pool tests and outposts--roughly in the order of decreasing risk in drilling. For a detailed discussion of exploratory drilling and definition of the above categories, see Lahee (11).

[3] Directionality indicates the probability of finding the product intended to be discovered. This concept will be discussed in greater detail in later pages.

is determined in terms of success or failure, and the size of the discovery in the case of success. The resultant new discoveries may be further added to by extensions and revisions.[1] Thus in order to study the response of changes in reserves to economic incentives, the various links in the exploratory process have to be analyzed in detail. It is proposed to develop such a framework in this paper and use it to formulate an empirically testable model of the process.

Fisher's model of the exploratory process consists of three stages with estimating equations for total exploratory wells, success ratios and sizes of discoveries of gas per successful well. An estimate of discoveries is obtained by multiplying the estimated wells by the success ratios and sizes. His model employs the "inventory of prospects" hypothesis. According to this hypothesis, "for economic as well as for geological reasons, the small prospects considered by operators tend to be relatively certain and the big prospects relatively risky...Over time an inventory of undrilled prospects accumulates about which something is known but which are either too small or too uncertain to be worth drilling under current conditions...(This leaves) a residue consisting largely of small relatively certain prospects."[2] He argues that when the price of crude oil rises, riskier (and poorer) prospects are drilled which may result in a lower success ratio and most likely, smaller average size of discoveries. In addition, Fisher postulates that as economic incentives increase, the outflow from the inventory tends to be predominately of small relatively certain prospects. This latter effect tends to offset the former by

---

[1] For a definition of these terms see Reserves of Crude Oil, Natural Gas Liquids, and Natural Gas in United States by AGA/API/CPA.

[2] Fisher (5), p. 405.

increasing the fraction of productive wildcats in current year (and due to the depletion of inventory, decreasing in the following year). Fisher's main point seems to be that the drilling decisions determine both the number of wildcats drilled and observed average characterics of the prospects. The relationship is an interdependent one.[1]

In a later study, Erickson and Spann (ES) extend the basic Fisher framework to take into account joint production aspects of oil and gas (See (3), (4) and (19)). They retain the Fisher argument that lagged success ratios (the ratio of productive to total wildcats) correlate positively with average discovery sizes and negatively with success ratios in the current year. In addition, they incorporate into their analysis certain characteristics of the industry--the accumulation of large size gas prospects in the earlier post-war years yielding disparate inventory conditions for oil and gas,as well as the occurrence of large gas prospects in association with large oil prospects. They derive conditions to be satisfied by the cross-price elasticities in their model, and evaluate their empirical results in the light of these conditions.

Khazzoom (8,9) visualizes the discovery of natural gas reserves as a response to a signal (gas or oil price) through a "black box" which includes such factors as the number of wildcats drilled, success in drilling and directionality. He concedes that the price signal triggers the drilling decision which determines simultaneously the number of wildcats drilled, the success ratio, directionality and the average size of discoveries. However, he proceeds to estimate the output of the black

---

[1] ibid. p.8 and passim.

box directly in terms of the inputs and stresses the lagged distributive effects of prices rather than structural aspects of natural gas industry. This is somewhat consistent with the policy nature of his model.

The MacAvoy-Pindyck (MP) model (14) treats simultaneously the field market for reserves (gas producers dedicated new reserves to pipeline companies at the well head price) and the wholesale market for production (pipeline companies selling gas to public utilities and industrial users). The linking of the two markets for purposes of policy simulation is the innovation of the MP approach. On the field market side (the supply side), they study the exploratory process in two stages--exploratory drilling and average size of discoveries per exploratory well. This avoids the difficulties of modelling the success ratio, admittedly the weakest link the the Fisher-ES estimations.

The present discussion attempts to provide a unified framework that takes into account various aspects of the exploratory process such as the uncertainty in drilling, depletion of the resource base with time, changing economic incentives and the joint nature of exploration for oil and gas. The equilibrium in the market for oil and natural gas prospects is described and the individual explorer's decision process is analyzed in the light of the characteristics of such an equilibrium. The analysis leads us to expectations regarding the nature of economic relationships satisfied by the relevant economic variables in the oil and gas exploration process. The implied economic relationships have been estimated after suitable modifications to eliminate econometric biases and to make them compatible with the available empirical data.

The goal here has been to derive relationships that (a) naturally follow from the choice behavior of an individual driller, and (b) lead to parameter estimates free of the econometric biases that may have crept into some of the earlier estimations. Since the success ratio equations have been the weakest link in many of the early studies, special attention has been devoted to this segment. An effort has been made to state all the assumptions implicit in the derivation of the econometric relationships as well as problems of aggregation associated with the estimation relationships. The relative merits of the one-stage (Khazzoom), the two-stage (M-P) and three-stage (Fisher, E-S) can be evaluated in the light of this framework. The model developed can be used as part of a more comprehensive model such as the MP model to investigate the effects of alternative regulatory policies on natural gas supplies.

Theory of the Exploratory Process

This section develops a concept of equilibrium in the market for oil and gas prospects. It provides a unified framework where the theory of expectations of the explorers is refined to include the effects of uncertainty and depletion of resources.

The point of departure of the theory presented here is the assumption that explorers in the petroleum industry have a notion of their intent in drilling a particular prospect--for oil, gas or hydrocarbon. The amount of exploratory drilling in each of these categories depends on the relevant economic incentives, the existing inventory of prospects and the attitudes of prospectors towards risk. For instance, a rise

in the price of oil, ceteris paribus, might lead to an increase in drilling for oil. The explorers' expectations as to the returns and risks in each category of drilling are reflected in the relative numbers as well as the characteristics of the wells drilled under each intended category (for oil, gas or hydrocarbons). Explorers take into account the fact that a particular well may or may not lead to a success, and, if a success, could yield gas, oil or both. As a matter of fact, there is an entire range of possibilities open to them as to the prospects they could choose to drill, corresponding to different relative proportions of oil and gas expected. The explorers look at the uncertain returns from drilling in each category, and compute the odds of various outcomes and the returns in each outcome. They choose the "package" that maximizes their expected utility (by trading off risk for return).

At any point in time there is an inventory of undrilled prospects about which some information is available. The prospects vary in size, probability of success, probability of finding oil or gas, and cost of exploration. The utility maximizing behavior on the part of the risk-averse explorers leads to an efficient inventory frontier.[1,2] Thus a prospect that yields a lower expected return for the same level of risk as another prospect is disqualified; conversely, a prospect

---

[1]The present approach is in marked contrast to the approach of Fisher and ES, who regard small, relatively certain prospects to be preferred to large uncertain prospects. The possibility that larger size may compensate an explorer for larger risk is not considered by them.

[2]Strictly speaking, the above exposition does not assume the perceptions regarding returns and risks of different prospects to be uniform across all individual explorers. Each individual might well have his own subjective efficient frontier of prospects. We will at a later stage, find it convenient to assume that expectations are nearly homogeneous across individuals. We will then be able to postulate a single common efficient frontier and model the deviation of an individual's perception of the frontier from the common frontier as a random "error."

that has a higher level of risk for the same expected return as another available prospect is rejected. The efficient inventory frontier includes small, relatively certain prospects as well as large but less certain prospects. The particular prospect chosen by an individual explorer depends on his utility function, in particular, his attitude to risk. The more risk- averse the explorer, the more likely it is that he will choose prospects that are expected to yield small but relatively certain returns.

The drillers' choice problem may be characterized as the choice between the three alternatives--drilling for oil, drilling for gas and not drilling.[1,2] The choice between the three alternatives is determined by the maximization of expected utility: the alternative chosen is the one yielding the highest utility to the driller after trading off risk against return.[3] An operational rule for the decision may be to pick a desired level of variance (and higher moments) and choose the alternative that maximizes the mean return.

---

[1]Exploratory activity where the primary motivation is the expectation of finding a gas deposit is referred to as "drilling for gas." "Drilling for oil" is defined similarly. Further, prospects explored under "drilling for gas" and "drilling for oil" are termed "gas prospects" and "oil prospects" respectively.

[2]Henceforth, our discussion will center only on oil and gas prospects since there is evidence to show that drilling for hydrocarbons (unassigned prospects)form a very small proportion of total exploratory drilling. (See Khazzoom (8), p. 33.) To the extent that relevant economic incentives and geological variables for hydrocarbon drilling are included in the final functional forms derived, drilling for hydrocarbons is accounted for in the model but not explicitly considered.

[3]In the mean-variance context, the optimizing behavior of the individual simply yields the condition that the marginal rate of transformation between mean return and variance of return is equal to the corresponding marginal rate of substitution implied by his utility function, i.e.,
(continued on next page)

We now proceed to characterize the "efficient inventory frontier" in greater detail. The analysis of the effect of uncertainty is carried out in mean-variance terms, i.e., it is assumed that the first two moments of the distribution are adequate to characterize the utility of a particular prospect to the driller.[1]

Let $\tilde{G}^j$ and $\tilde{O}^j$ be the random variables representing the anticipated sizes of gas and oil discoveries from the j^{th} prospect under consideration. Let:[2,3]

$$\mu_G^j = E(\tilde{G}^j)$$

$$\mu_O^j = E(\tilde{O}^j)$$

$$V_G^j = Var(\tilde{G}^j)$$

$$V_O^j = Var(\tilde{O}^j)$$

$$V_{GO}^j = Cov(\tilde{G}^j, \tilde{O}^j).$$

(Continued from page 7)

$$[d\mu_R/d\sigma_R^2]_{\Phi_E} = [d\mu_R/d\sigma_R^2]_{U_i}$$

where $\Phi_E = 0$ represents the efficient frontier of prospects (assumed to be continuous), μ_R is the expected dollar return, σ_R^2 is the variance of dollar return and $U_i()$ is the function of the individual's utility function.

[1] There are well-known drawbacks of such an assumption since it is valid only with a quadratic utility function, when the returns are normally distributed or when the distributions of returns are "compact." See Tobin (20) or Samuelson (16). All these explanations are implausible in a theoretical sense. However, the first two moments of the distributions of returns are a good approximation in several practical situations, e.g., portfolio theory. The exposition is considerably simplified by restricting it to two moments.

[2] It seems reasonable to expect that V_{GO}^j will be positive in the case of fields where gas discoveries are "associated" (with oil) and zero or even negative in the case of "non-associated" discoveries.

[3] (See next page)

Clearly, other things being equal, higher values of μ_G^j and μ_O^j are more desirable while higher values of V_G^j, V_O^j and V_{GO}^j are less desirable.[1] Each prospect can be represented by a point in the five-dimensional space with μ_G, μ_O, V_G, V_O and V_{GO} as the axes. We shall refer to points that have the highest values of (μ_G^j, μ_O^j) for a given combination of the values of (V_G^j, V_O^j, V_{GO}^j) as the "efficient points with respect to the currently available set of prospects, and the locus of all such efficient points to as the efficient inventory frontier, and is a five-demensional surface represented by

$$\Phi_E(\mu_G, \mu_O, V_G, V_O, V_{GO}) = I$$

where I is the level of inputs into the exploratory activity.

A second five-dimensional surface may be defined to represent the boundary separating prospects that are worth drilling in the opinion of explorers from those that are not worth drilling. This is called

---

[3](from p. 9)
μ_G is measured in Mcf , μ_O in barrels, V_G is measured in $(\text{Mcf})^2$, V_O in$(\text{barrels})^2$, and V_{GO} in Mcf barrels.

[1]In the case where there is no uncertainty, the five-dimensional efficient frontier reduces to a two-dimensional frontier with the axes representing μ_G and μ_O. This is analogous to the production possibility curve postulated by Erickson and Spann (3).

the "critical profitability frontier" and is represented by

$$\Phi_P(\mu_G,\mu_O,V_G,V_O,V_{GO}) = I$$

The prospects lying between the efficient frontier and the critical profitability frontier are those that the explorers would like to drill, the first prospects drilled always being those on the efficient frontier. As drilling goes on, the efficient inventory frontier shifts in position and moves closer to the critical profitability frontier. The process of drilling stops only when the two frontiers conicide. The area between the two frontiers determines the amount of drilling while the "shape" of the area determines the type of prospects drilled.

The efficient inventory frontier thus represents the "best" currently available prospects and is the result of all the past actions of explorers. It is unaffected by changes in economic incentives but does shift as information on new prospects (or new information on old prospects) becomes available and as depletion of prospects takes place. On the other hand, the critical profitability frontier shifts in response to changes in economic incentives, but not due to new geological information or depletion of prospects.

Assuming a wide range of prospects covering a spectrum of values of $\mu_G,\mu_O,V_G,V_O,V_{GO}$ is available, the efficient frontier as well as the critical profitability frontier may be expected to be nearly continuous and "smooth," satisfying the usual mathematical regularity conditions.

Note in the above context, that the so-called extensive drilling[1] (drilling of large, relatively uncertain prospects) will correspond

---

[1] See MacAvoy and Pindyck (14).

to drilling in the high μ_G, μ_O region of the efficient frontier while the reverse is true of intensive drilling (drilling of small, relatively certain prospects).

The five-dimensional efficient frontier $\Phi_E()$ is specified independently of information on the relative prices of gas and oil. In the light of price information, explorers can evaluate the worthiness of prospects in terms of their dollar return.[1] For purposes of exposition the efficient inventory frontier and the critical profitability frontier can be mapped on to two dimensions with means and variances of dollar returns as the two axes. To do this, the random variable estimated return from the j^{th} prospect is computed as:

$$\tilde{R}^j = PG.\tilde{G}^j + PO.\tilde{O}^j - C^j$$

where PG and PO are prices of natural gas and oil respectively and C^j is the average total cost of exploring the j^{th} prospect further. Then

$$\mu_R^j = E(\tilde{R}^j) = PG.\mu_G^j + PO.\mu_O^j - C^j$$

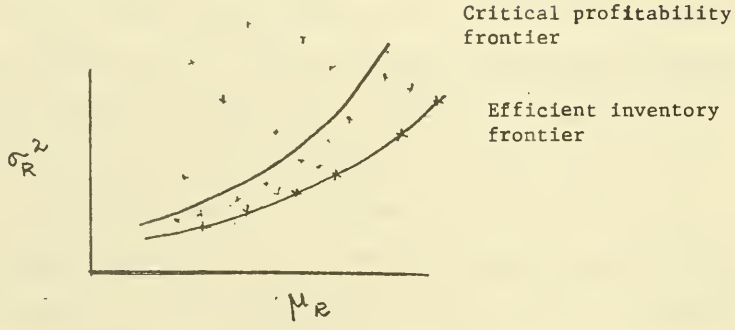
$$\sigma_R^{j\,2} = \text{Var}\,(\tilde{R}^j) = PG^2.V_G^j + PO^2.V_O^j + 2PG.PO.V_{GO}^j$$

assuming that exploration costs are known with certainty. Under the assumption of risk-averse explorers, the efficient inventory frontier

---

[1] The dollar returns from a discovery may be spread over a number of years. Strictly speaking, therefore, the explorers look at the discounted present value of these returns. In the absence of specific information on the distribution of cash flows from discoveries over time, we ignore this effect and assume that $\tilde{R}^j$ indicates the return discounted appropriately. One way to take this effect partially in to consideration is to include the interest rate as an independent variable in the model (See Spann and Erickson (19)).

as well as the critical profitability frontier will be upward sloping as sketched below:



Different points along the efficient inventory frontier might correspond to different anticipated gas to oil ratios. As a matter of fact, some of the points would represent "gas prospects" while others would be "oil prospects" (For definition of gas and oil prospects, refer to footnote 1, p. 7). The choice between drilling for gas and drilling for oil depends on how the individual explorer's risk preferences compare with the risk-return trade off available from the gas or oil prospect concerned.

We can now consider the effects of shifts in different economic and geological variables on the positions of the efficient inventory frontier and/or critical profitability frontier. A shift in the economic incentives will leave the efficient inventory frontier unchanged, but the critical profitability frontier is likely to shift. For instance, a hike in the price of gas may justify certain prospects hitherto considered uneconomical. A change in the tastes and perceptions of explorers

is likely to cause a movement along the efficient frontier of different drillers rather than a shift of the frontier--if drillers become more sensitive to risk, they will shift towards small, more certain prospects. Technological advances tend to shift the critical profitability frontier in a direction favorable to the drillers ("upward"). Changes in the composition of the set of available prospects occurring due to the depletion of resources over time are reflected as upward shifts of the efficient inventory frontier but do not shift the critical profitability frontier.[1] Unanticipated new information concerning prospects is assumed to arrive randomly and hence its effects on the efficient frontier and the critical profitability frontier are properly treated as random "error."[2]

Note that the above discussion of means and variances does not distinguish the outcome of the drilling process between the probability of success and the size of discovery conditional on a success. However,

---

[1] Depletion effects are considered in this approach as long-run effects as opposed to the Fisher treatment as a one-period lagged effect. Thus, the concept of long-run, continuing depletion of prospects replaces Fisher's description of depletion of inventory of prospects.

[2] This description of the drilling process in terms of shifts in the efficient inventory frontier and the critical profitability frontier is in marked contrast to the Fisher--ES view that price increases would lead to shifts towards smaller, but more certain prospects. There is reason to believe that there is always a predominant proportion of small, certain prospects because of natural geological reasons. But, if the question is whether there is a shift in the composition of prospects selected by drillers in response to a price hike, the answer has to be 'no.' The stochastic return $\tilde{R}^j$ is altered by the same percentage in the case of all prospects, for any given change in the prices PG and PO. The type of prospect (large, risky or small, certain) an individual would choose depends on the nature of his "indifference curves" between risk and return as well as the distribution of prospects along the frontier. In our view, Fisher's description involves "double counting" of the risk effect. As a matter of fact, there may be some "economies of scale" effects associated with large sized prospects, which may counteract the price effect described by Fisher-ES.

under certain assumptions, we may be able to postulate some hypotheses regarding the joint distribution of size of discovery and success ratios. For instance, assuming that, on the whole the average, expectations are realized, drilling has high directionality, and negligible drilling for hydrocarbons (unassigned) takes place[1], high success ratios can be expected to be associated with small sizes of discovery and vice versa. This matter will be dealt with in greater detail in a later section.

In actuality, we only have data on total wells drilled, size of discovery, and success ratio (by production district). Thus, the observed characteristics of the prospects drilled correspond to the weighted averages of all points on the efficient-profitable band (the region between the efficient inventory frontier and the critical profitability frontier).

Erickson and Spann (3, 19) discuss in considerable detail the implications of the fact that exploration costs are joint. One of the conclusions that follows from their analysis is that the cross price elasticities of oil and gas must be equal (or proportional). They use this result in evaluating the quality of their results. (In their latest estimations [19], these constraints are actually imposed on the values of the parameters.) The conditions derived therein are appropriate in a world where the chances of success and sizes are known with certainty. Incorporation of the role of uncertainty into the analysis changes the

---

[1] In addition, the "free-rider" problem is assumed away--the results of exploratory drilling may reflect points off the efficient frontier to the extent that an otherwise uneconomical gas prospect is drilled since it occurs with an efficient oil prospect under high directionality, this problem will not be significant.

picture considerably. Cross-elasticity conditions of the type

$d\mu_G/dP_O = d\mu_O/dPG$ (or modifications thereof) need not be satisfied any more.

Both prices and costs of exploration enter into the computation of dollar returns, $\tilde{R}^j$. While prices are known separately for oil and gas, the same is not true of the costs. To the extent that the driller can choose any one of a spectrum of gas-to-oil ratios (as per his a priori estimates), the costs of finding oil or gas will be treated by him as joint costs, except in the extreme cases where he choses a gas-to-oil ratio of 0 or 1 [1,2,3]. Thus, for a particular level of inputs (and costs), a trade off exists between anticipated returns from oil and gas discoveries.

To see this, we consider the case of joint production of gas and oil in a mean-variance world and investigate the effects of changes in prices. The production possibility frontier faced by the driller is a five-dimensional surface for each level of joint exploratory costs. (Costs and input quantities are analogous if perfect competition is assumed in the factor markets.) The joint production possibility frontier (the efficient inventory frontier) is defined by

$$I = \Phi_E(\mu_G, \mu_O, V_G, V_O, V_{GO})$$

---

[1] The concept of joint exploratory costs has been vigorously advocated by ES (3,19) and by Fisher (5) and MP (14) as well.

[2] Thus in Marshallian terminology, from the point of view of exploratory costs, oil and gas are joint products with variable proportions. This is not inconsistent with the possibility that the actual production costs may be joint costs in fixed proportions, once a discovery has been made.

[3] In the special case where an explorer opts to search solely for gas (or solely for oil) any gas (or oil) he may obtain is a by-product rather than a joint product.

where I is the quantity of the input,[1] μ_G and μ_O are the mean sizes of gas and oil discoveries, V_G and V_O are the variance of the gas and oil discoveries, and V_{GO} is the covariance between oil and gas sizes, in the same prospect. To be an efficient point on the frontier, a high variance of gas size, V_G, may be accompanied with one or more of a combination of higher mean oil size, higher gas size, lower V_O and lower V_{GO}. Assuming costs to be known with certainty, the problem facing the explorer is:

$$\text{Max } U(\mu_R, \sigma_R^2)$$

subject to:

$$C = rF(\mu_G, \mu_O, V_G, V_O, V_{GO})$$

$$\mu_R = PG.\mu_G + PO.\mu_O - C$$

$$\sigma_R^2 = PG^2.V_G + PO^2.V_O = 2PG.PO.V_{GO}$$

where r is the factor cost, PG and PO are the prices of gas and oil.

Substituting back in the utility function, we have

$$\text{Max } U\left[PG.\mu_G + PO.\mu_O - F(\mu_G, \mu_O, V_G, V_O, V_{GO}) \;,\; PG^2.V_G + PO^2.V_O + 2PG.PO.V_{GO} \right]$$

The first order necessary conditions for the optimum are

$$\frac{\partial U}{\partial \mu_G} = \frac{\partial U}{\partial \mu_R}[PG - rF_1] = 0$$

$$\frac{\partial U}{\partial \mu_O} = \frac{\partial U}{\partial \mu_R}[PO - rF_2] = 0$$

$$\frac{\partial U}{\partial V} = \frac{\partial U}{\partial \mu_R} rF_3 + \frac{\partial U}{\partial (\sigma_R^2)} - PG^2 = 0$$

$$\frac{\partial U}{\partial V} = \frac{\partial U}{\partial \mu_R} rF_4 + \frac{\partial U}{\partial (\sigma_R^2)} \cdot PO^2 = 0$$

$$\frac{\partial U}{\partial V_{GO}} = \frac{\partial U}{\partial \mu_R} rF_5 + \frac{\partial U}{\partial (\sigma_R^2)} \cdot PG.PO = 0$$

[1] The analysis can be easily extended to several inputs when there is perfect competition in the market for each input.

where F_1, F_2, F_3, F_4, and F_5 are the partial derivatives of F with respect to μ_G, μ_O, V_G, V_O, and V_{GO} respectively. From the first two conditions,

$$PG/PO = F_1/F_2$$

Similarly, from the third and fourth conditions,

$$PG^2/PO^2 = F_3/F_4$$

Let F_{ij} represent the respective second partial derivatives and λ, the degree of absolute risk aversion defined by $\lambda = \dfrac{\partial U/\partial \sigma_R^2}{\partial U/\partial \mu_R}$ and differentiating totally, the five conditions above,[1]

$$dPG = r[F_{11}d\mu_G + F_{12}d\mu_O + F_{13}dV_G + F_{14}dV_O + F_{15}dV_{GO}]$$

$$dPO = r[F_{21}d\mu_G + F_{22}d\mu_O + F_{23}dV_G + F_{24}dV_O + F_{25}dV_{GO}]$$

$$2PG.dPG = r\lambda[F_{31}d\mu_G + F_{32}d\mu_O + F_{33}dV_G + F_{34}dV_O + F_{35}dV_{GO}]$$

$$2PO.dPO = r\lambda[F_{41}d\mu_G + F_{42}d\mu_O + F_{43}dV_G + F_{44}dV_O + F_{45}dV_{GO}]$$

$$2PG.dPO + 2PO.dPG = r\lambda[F_{51}d\mu_G + F_{52}d\mu_O + F_{53}dV_G + F_{54}dV_O + F_{55}dV_{GO}]$$

Assuming the smoothness properties $F_{ij} = F_{ji}$, it can be shown from these equations that the following condition must be satisfied by the total

---

[1] λ is not a constant except in the case of constant absolute risk aversion. Utility functions consistent with λ include the family of exponential utility functions generally. See Arrow (1) for a discussion.

derivatives with respect to product prices:

$$\frac{d\mu_G}{dPO} + 2PG\frac{dV_G}{dPO} + 2PO\frac{dV_{GO}}{dPO} = \frac{d\mu_O}{dPG} + 2PO\frac{dV_O}{dPG} + 2PG\frac{dV_{GO}}{dPO}$$

This is the condition that must be satisfied by shifts in the critical profitability frontier in response to changes in product prices, for a given level of inputs. This condition is analogous to the "cross derivative" condition arrived at by ES [(3), (19)] for the case when only mean returns are taken into account. In the present analysis, changes in price affect not only the mean but also the risk characteristic represented by variances and the covariance of returns from exploration.[1]

The Individual's Drilling Decision

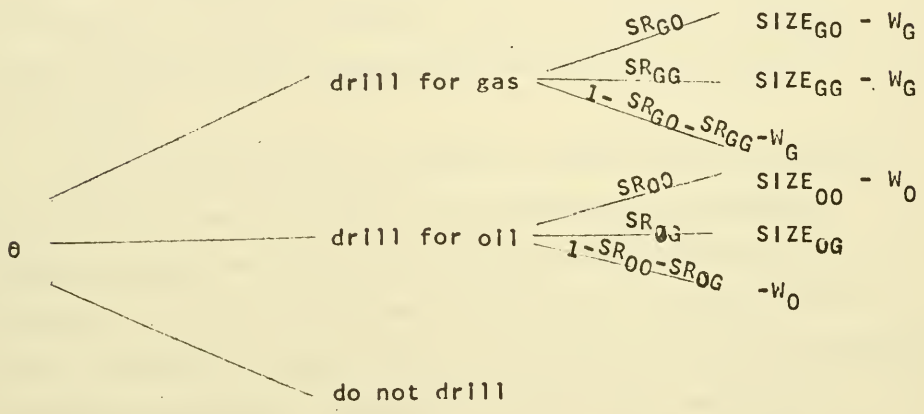
The previous sections describe the environment confronted by an individual driller in terms of the efficient inventory frontier and the critical profitability frontier. The environment is at the same time the result of and the stimulus for, the actions of the individual driller. The individual driller looks at the currently available efficient set of prospects and makes his decision regarding whether or not to drill and if the decision is to drill, which prospect to drill. The type of drilling may be broadly classified as "drilling for oil" or "drilling for gas." (As stated earlier, we assume that unassigned drilling is negligible, but loosening this assumption causes no major conceptual problems.) The total number of exploratory wells drilled as well as the relative fraction of wells intended for gas and oil are outcomes of the collective

---

[1] It should be noted that these conditions have to hold only for the total derivatives with respect to prices. This must be kept in mind when interpreting the signs of coefficients in a least squares regression; this can be done only if none of the right hand side variables is a function of PG or PO.

actions of individual explorers. This motivates the following derivation of the functional forms for total wells and their relative fractions based on a consideration of the individual's decision process.

The decision process of the individual driller may be characterized by the following decision tree:



The SR variables stand for success ratios in each category and the SIZE variables stand for sizes of discovery conditional on success. The subscripts OG,GG, GO and OO refer respectively to discovering gas when oil was intended, gas when gas was intended, oil when gas was intended and oil when oil was intended. Thus, SR_{OO} = probability to discovering oil if an oil prospect is drilled, $SIZE_{OO}$ = random variable representing the size of oil discovered if the above occurs. We shall henceforth refer to SR_{OO} and SR_{GG} as "intended" success ratios for oil and gas and $SIZE_{OO}$ and $SIZE_{GG}$ as the "intended" sizes. W_O and W_G are the costs of exploratory drilling per oil and gas well respectively.

The stochastic returns per well $\tilde{R}_G$ and $\tilde{R}_O$ from drilling for oil and gas respectively may be expressed in terms of prices, success ratios and sizes, as

$$\tilde{R}_G = (SR_{GO})\ (SIZE_{GO})\ PO + (SR_{GG})\ (SIZE_{GG})\ PG - W_G$$

$$\tilde{R}_O = (SR_{OO})\ (SIZE_{OO})\ PO + (SR_{OG})\ (SIZE_{OG})\ PG - W_O$$

The individual driller has subjective estimates for each of the above variables. In making his drilling decision, the driller can confine his attention to the set of prospects lying on the efficient inventory frontier because all other prospects are dominated by some prospect on the frontier.

Since the characteristics (including the various expected fractions of success and sizes of discovery) of the prospects on the efficient inventory frontier are influenced by economic as well as geological variables, we express anticipations of success ratios and sizes of discovery as functions of these variables. The effects of depletion of gas and oil reserves are considered by including the variables DEP_G and DEP_O (to be defined later) as the relevant indices. The economic incentives are represented by the prices PG and PO and the average exploration cost per well, ATC. This leads to the following functional forms for each of the SR and SIZE variables:[1]

---

[1]The estimation of the equations will be done using pooled cross-section time series data. For the time being, district subscripts are dropped for expositional convenience and the analysis is carried out ignoring geological differences across regions. A later section describes how the geological differences are taken into account in the estimation.

$$SR = \Psi_1(PG, PO, ATC, DEP_G, DEP_O)$$

$$SIZE = \Psi_2(PG, PO, ATC, DEP_G, DEP_O)$$

The dollar returns $\tilde{R}_G$ for the k^{th} gas prospect on the efficient frontier can be written as

$$\tilde{R}_G^k = R_G^k(PG, PO, ATC, DEP_G, DEP_O)$$

Similarly for the 1^{th} oil prospect lying on the efficient frontier

$$\tilde{R}_O^{\ell 1} = \tilde{R}_O^{\ell 1}(PG, PO, ATC, DEP_G, DEP_O)$$

Let S_G denote the set of subscripts of all currently efficient gas prospects and S_O that of all efficient oil prospects. Let $U_i(x)$ represent the utility to the ith driller of x dollars of returns from exploratory activity. We shall arbitrarily assign a utility of zero to the status quo. This is the utility associated with deciding not to drill any further.

The i^{th} driller will choose to drill one of the prospects in the inventory (as a matter of fact, the prospects on the current efficient frontier) if and only if[1]

$$\text{Max}_{k \epsilon S_G,\ \ell \epsilon S_O} \left[U_i(\tilde{R}_G^k),\ U_i(\tilde{R}_O^{\ell 1}) \geq 0\right]$$

In the case when this condition is satisfied, he will drill for gas if

$$\text{Max}_{k \epsilon S_G} \left(U_i(\tilde{R}_G^k)\right] > \text{Max}_{\ell \epsilon S_O} \left[U_i(\tilde{R}_O^{\ell 1})\right]$$

and for oil otherwise. He will be indifferent between drilling for gas and drilling for oil (and drilling for hydrocarbons) in case of equality.

Let X_i be the set of attributes that influence the i^{th} individual's estimates of $\tilde{R}^*_G$ and $\tilde{R}^*_O$. The vector X_i includes the values of the economic and geological variables as observed by the i^{th} individual. The vector X of the same attributes, but as observed by us, cannot capture the variations in perceptions of the alternatives among individuals. An explicit way of stating this property would be write $U_i(X_i)$ as

$$U_i(X_i) = U_i(X,\eta_i)$$

where η_i is an unobserved vector associated with the i^{th} individual's perception. In addition, the utility functions of different individuals are not identical although they are assumed to belong to the same family (e.g., the logarithmic or exponential utility functions). Even with the same set of perceptions and the same decision rules, different individuals may be led to different decisions. In order to characterize the collective impact of several individual explorers acting simultaneously, we postulate a reference utility function for dollar returns. The i^{th} individual's utility differs from $\bar{U}(X)$ which may be thought of as the "average typical" explorer's utility function by a random deviation G_i:

$$U_i(X_i) = U_i(X,\eta_i) = \bar{U}(X) + \nu_i$$

The ν_i terms are outcomes of the random variables $\tilde{\nu}$ and represent unobserved deviations capturing the characteristics of the individual that cannot be measured, and are assumed to be randomly distributed across

individuals.[1]

The i^{th} individual's decision rules can now be expressed in terms of the "average typical" explorer's utilities as:

Drill an exploratory well if

$$\underset{\substack{k\epsilon S_G \\ l\epsilon S_O}}{\text{Max}} \left\{ \bar{U}(X_G^k) + \nu_{Gi}^k,\ \bar{U}(X_O^l) + \nu_{Oi}^l \right\} \geq 0$$

ν_{Gi}^k and ν_{Oi}^l are the deviations for the k^{th} gas prospect and the l^{th} oil prospect. These deviations are outcomes corresponding to the i^{th} individual of the random variables $\tilde{\nu}_G$ and $\tilde{\nu}_O$ respectively.

Given that a decision is made in favor of drilling, he will drill for gas if

$$\underset{k\epsilon S_G}{\text{Max}} \left\{ \bar{U}(X_G^k) + \nu_{Gi}^k \right\} \geq \underset{l\epsilon S_O}{\text{Max}} \left\{ \bar{U}(X_O^k) + \nu_{Oi}^k \right\}$$

and for oil otherwise.[2]

More operationally, in a mean-variance context, we can think of the individual explorer as having a notion of the level of risk (in the

---

[1] If all explorers had identical perceptions and utility functions, all exploration would be for either oil or all for gas, depending on the state of nature. However, individual tastes and perceptions vary and cause a distribution of drilling across gas and oil prospects, changing the relative frequencies of oil/gas prospects. There may be from time to time prospects on which new information, hitherto unavailable, becomes known. These may deviate considerably from the rest of the currently efficient prospects. However, these occur randomly and cannot be modeled explicitly except through the error term.

[2] The knife-edge case of "unassigned" drilling has been arbitrarily included under "drilling for gas." We shall continue to do so through the rest of this analysis. There is evidence [see (8)] to show that the amount of unassigned drilling is quite small.

form of variance of returns) he would be willing to accept;[1] and making his drilling decision by looking at the estimated mean returns of the different prospects. Let the level of variance acceptable to him be $\bar{V}$.[2] The explorer can confine his attention to the efficient gas prospect and efficient oil prospect, both of which are expected to yield random returns of variance $\bar{V}$. Let the average-typical explorer's estimates of the returns from this efficient gas prospect and the efficient oil prospect be $\tilde{R}_G^*$ and $\tilde{R}_O^*$ respectively.

Thus, for the average-typical man,

$$\text{Var}(\hat{R}_G^*) = \text{Var}(\hat{R}_O^*) = \bar{V}$$

He will drill an exploratory well if

$$\text{Min}\,[E(\hat{R}_G^*),\ E(\hat{R}_O^*)] \geq \bar{R}^*$$

where $\bar{R}^*$ is the critical level of the estimated mean return that he is willing to accept for the prescribed level of variance. Given that a decision is made to drill, he will drill for gas if

$$E(\hat{R}_G^*) > E(\hat{R}_O^*)$$

and for oil otherwise.

The decision rule for the i^{th} individual explorer is exactly the same as above except in that his estimates of the variances and means

---

[1] He will formulate this notion after taking into account information on the set of currently available prospects and the economic environment.

[2] Since each explorer accepts several prospects, the relevant "risk" associated with a particular prospect is the contribution of this prospect to the risk of the entire portfolio. In general, the contribution to risk of an individual prospect is not equal to the "own risk" of the prospect evaluated independently There is reason to believe that the returns from individual prospects do not exhibit strong correlations with each other. If, in fact, two or more prospects under consideration by an individual driller are not independent, the driller is likely to treat the whole "package" as a single prospect. The strategy he

(continued on next page)

of $\tilde{R}_G$ and $\tilde{R}_O$ differ from those of the "average-typical" individual's by deviations ε_{Gi} and ε_{Oi}, and his critical level of mean return $\bar{R}_i$ differs from $\bar{R}^*$ by a deviation ε_{Ri}. (ε_{Ri} are outcomes of the random variable $\tilde{\varepsilon}_R$.) Thus, when a large number of explorers are operating simultaneously, the probability that an exploratory well is drilled[1]

$$= \Pr[\text{Min}\{E(\tilde{R}_G^*) + \tilde{\varepsilon}_G, E(\tilde{R}_O^*) + \tilde{\varepsilon}_O\} \leq R^* + \varepsilon_R]$$

$$= \Pr[\text{Min}\{E(\tilde{R}_G^*), E(\tilde{R}_O^*)\} - R^* \leq \tilde{\delta}]$$

where $\tilde{\delta}$ is a random variable that depends on $\tilde{\varepsilon}_G$, $\tilde{\varepsilon}_O$, and $\tilde{\varepsilon}_R$.

Since $\tilde{R}_G$ and $\tilde{R}_O$ are functions of the vector of economic and geological variables, so are their expectations. The total number of exploratory wells drilled depends directly on the above probability and can therefore be expressed in the functional form:

$$WXT = \psi_3(PG, PO, ATC, DEP_G, DEP_O, RISK)$$

where the variable RISK is an index of the chosen level of $\bar{V}$. The last

---

2 (continued from page 24)
would follow is probably to drill only one out of the package to start with and consider drilling others only after getting the benefit of information from the first. Thus, at any point in time, only one prospect out of the "package" is under his serious consideration for drilling. Assuming the returns are independent of all other prospects in the portfolio, the risk of an individual prospect can be evaluated independent of all other prospects in the portfolio. Thus, in a mean-variance context, the contribution to risk of the i^{th} would be the own variance σ^2_{Ri} since the covariance terms are all zero.

If the driller intends to follow the operational rule of setting a desired mean level $\bar{R}$ of dollar returns from the portfolio and minimizing the corresponding variance, he can do so without paying attention to the so-called portfolio effect, simply by successively picking prospects that have minimum variance for a given level of mean.

[1] In this mean-variance analysis, the random variables $\tilde{\varepsilon}_G$ and $\tilde{\varepsilon}_O$ respectively indicate the deviations between the average individual's estimate of mean returns $E(\tilde{R}_O^*)$ and any particular individual's estimate of mean returns $E(\tilde{R}_O^*)$ and any particular individual's estimates. For the i^{th} individual the outcomes of $\tilde{\varepsilon}_G$ and $\tilde{\varepsilon}_O$ are respectively ε_{Gi} and ε_{Oi}. Similarly the critical level of mean returns $\bar{R}^*$ varies by the random variable $\tilde{\varepsilon}_R$ (which takes the value ε_{Ri} for the i^{th} individual).

variable on the right hand side is a new addition to the earlier list of variables. It enters because we used the level of risk chosen by as a parameter in the above discussion. Higher prevailing levels of risk aversion among explorers will lead to lower levels of exploratory activity. Based upon the earlier description of the equilibrium and the efficient frontiers, we expect a positive sign for the coefficients of the prices whereas ATCD, DEP_G, DEP_O, and RISK are expected to enter with negative coefficients.

Given that a decision has been made to drill, the probability that it will be drilled with intent for gas is given by

$$p = Pr[E(\tilde{R}_O^*) - E(\tilde{R}_G^*) < \tilde{\varepsilon}_G - \tilde{\varepsilon}_O]$$
$$= g[E(\tilde{R}_O^*) - E(\tilde{R}_G^*)]$$

where g is the cumulative distribution function of the stochastic term $(\tilde{\varepsilon}_G - \tilde{\varepsilon}_O)$.[1] The probability that it will be drilled with intent for oil is

$$p = (1 - g)$$

The specific probability function of g is obtained as a result of sampling from a population of utility-maximizing explorers whose deviations from the reference are distributed according to the distribution function g.

---

[1] It can be shown that if $\tilde{\varepsilon}_G$ and $\tilde{\varepsilon}_O$ are statistically independent with identical exponential distribution

$$\text{Prob}\,[\varepsilon_{G,O} \leq w] = e^{-e^{-w}}, \text{ then Prob}\left\{\varepsilon_G - \varepsilon_o \leq t\right\} = \frac{1}{1 + e^{-t}},$$

which is the cumulative logistic distribution. This will form some of the motivation behind the logistic forms presented later.

Since the probability p is governed by the relative values of the expectations of the returns from oil and gas, it can be expressed as a function of the relative price PG/PO, cost ATC and the depletion and risk variables for gas and oil. Hence

$$p = g[\Psi_4(PG/PO, ATC, DEP_G, DEP_O, RISK_G, RISK_O)]$$

where $RISK_G$ and $RISK_O$ are indices of risk associated with "gas wells" and "oil wells" respectively and are defined as the variances of the returns from "gas wells" and "oil wells" respectively.

If WXTG are the number of exploratory wells drilled with intent for gas, WXTO those drilled for oil, and WXT is the total exploratory wells drilled, the fractions WXTG/WXT and WXTO/WXT become estimates of the probabilities p and g respectively. Thus the fraction of total wells drilled for gas can be written as:[1]

---

[1] One may view the outcome WXTO and WXTG as outcomes of random variables. An example of this approach is that of Kaufman (7) who views WXTO and WXTG as Poisson distributed variables whose means μ_1 and μ_2 are determined as functions of economic and geological variables such as those that appear on the right hand sides above. He obtains reasonable fits for the description of the explanatory process as a Poisson distribution without replacement for individual fields based on data for Alberta Province, Canada. Assuming that the means μ_1 and μ_2 are the same for all fields in a district (the only reasonable assumption possible without detailed data on each field), deviations from the mean of each individual production district will be the sum of a number of variables of the form $f_p(\mu_i) - \mu_i$ where f_p is the Poisson density function. If each prospect's results are independent of the next, and the number of fields is large, the distribution of the deviation term to be approximately normal when the number of fields is large. We can thus expect the least squares estimates of the WXT equation to be close to the maximum liklihood estimates.

$$\frac{WXTG}{WXT} = \Phi_G(PG/PO, ATC, DEP_G, DEP_O, RISK_G, RISK_O)$$

Similarly,

$$\frac{WXTO}{WXT} = 1 - \Phi_G() = \Phi_O\ (PG/PO, ATC, DEP_G, DEP_O, RISK_G, RISK_O)$$

We expect the price ratio PG/PO to have a positive effect on Φ_G and a negative effect on Φ_O. The depletion variable DEP_G and risk variable $RISK_G$ should have a negative effect on Φ_G (and hence a positive effect on Φ_O). The reverse would be true of the variables DEP_O and $RISK_O$.

Intended Success Ratios and Sizes of Discoveries

The explorers also have a notion of their intended return per well besides knowing whether they are drilling for gas or oil. These intentions or expectations are formed on the basis of information from the prevailing equilibrium in the market for gas and oil prospects and the explorers' own attitudes towards risk (degree of risk averseness).

We will find it convenient to describe the characteristics of these anticipated returns in terms of intended success ratios and intended average sizes of discoveries, conditional on a success. This breakdown is convenient for several reasons:

Firstly, a producer who is engaged in exploratory activity has, at any time, a choice as to whether any increases in his drilling activity will be extensive or intensive. (See Fisher (5), and MacAvoy-Pindyck (14)). If he follows the extensive mode of drilling, relatively small numbers of wells are drilled, but they usually go beyond the frontier of recent discoveries to open up new geographical locations or new strata in old locations. If on the other hand he follows the intensive mode of drilling, many small wells are drilled in an area that had already proven

itself to be a probable source of gas discovery. Thus extensive drilling is associated with relatively small probabilities of success, but conditional on success, size of discovery is likely to be large. In intensive drilling, probability of success is large, but sizes are likely to be small and narrowly distributed. In the population of drillers, there are some who are looking for large success ratios, but are satisfied with small sizes, while others looking for large sizes of discovery are prepared to accept the possibility of very small success ratios. By separating the success ratio and size aspects, we can get better information on whether the intensive mode or the extensive mode is currently the dominant one among explorers. This information can be used to simulate the effects of changing economic incentives (or other influences). The amount of confidence we like to place on the structural forms of the model will depend on the internal consistency of the coefficients estimated in the success ratio and size equations as well as how well the results of the simulations of the past are borne out by reality. Such a possibility of checking the validity of the structure is an important advantage if the model is to be used for studying policy alternatives.

Secondly, separating the two effects will avoid some aggregation problems. In a world where some people prefer to drill extensively and some intensively, estimating discoveries per exploratory well rather than per successful well involves superimposing two probabilities of distribution of considerably different characteristics. Whether the reduced aggregation biases lead to a better forecast of discoveries or not, depends on the net results of two conflicting factors: increased accuracy in the individual structural estimates and the greater number

of equations to be estimated.

A second kind of aggregation bias too is avoided by using the 3-stage description of the process. Most of the non-associated discoveries of gas are obtained from gas wells while associated gas comes mostly from oil wells. If successful gas wells and successful oil wells have significantly different characteristics (for instance, the average size of a successful gas well may be affected differently by economic incentives than by average size of successful oil well), it will be more appropriate to define the size of non-associated discoveries on a per successful gas well basis and size of associated discoveries on a per successful oil well basis.

Finally, separating the success ratio and size aspects also offers certain econometric conveniences: a usual problem faced in time series, cross-section estimations is that the form assumed may simply widely varying elasticities across cross-sections, because of widely varying levels of the dependent variable. This problem does not arise from the success ratio equations since they vary in a narrow range across regions, and so may allow superior estimation for this stage of the estimation. Also, the existence of depletion effects for gas and oil can be investigated through their influence on the success ratios. Further, in case a non-linear functional form is appropriate for the anticipated return from an exploratory well, splitting the return into the success ratio and size will allow greater flexibility in capturing the non-linearities.

Let SR^{*}_{GG} and SR^{*}_{OO} represent the <u>intended</u> success ratios for gas and oil drilling respectively, ie., anticipated success ratios for gas in wells intended for gas and for oil intended for oil. Similarly,

let $SIZE^*_{GG}$ and $SIZE^*_{OO}$ be the intended average sizes. These intentions are formed in the light of information about the set of prospects available to them and prevailing conditions of equilibrium in the market for prospects. When large depletions of reserves occur, the efficient inventory frontier shifts up, the available success ratios and sizes will both be lower, which in turn lowers the driller's anticipated value of SR and SIZE variables. We, therefore, expect DEP_G to have a negative effect on SR^*_G and $SIZE^*_G$ while DEP_O should have a negative effect on SR^*_O and $SIZE^*_O$. The directions of shifts in the SR and SIZE variables in response to changes in economic incentives are a function of whether the intensive or extensive mode of drilling is the predominant one. If the extensive mode of drilling is predominant in exploration for gas we expect a hike in price of gas (or a fall in costs of gas exploration) to have a negative effect on the intended success ratio for gas SR^*_{GG} and a positive effect on the corresponding intended size $SIZE^*_{GG}$. The reverse would be true if the drilling is predominantly intensive. The argument is similar for SR^*_{OO} and $SIZE^*_{OO}$. The risk variables should have a positive effect on both pairs of intended sizes and success ratios. Higher levels or risks in gas exploration would lead investors to demand higher mean returns in the form of greater success ratios and average sizes of discovery. Similarly higher levels of risk in oil exploration should be associated with higher levels of SR^*_{OO}

and $SIZE^*_{00}$. Let the following functional forms represent the intended (anticipated) success ratios and sizes:

$$SR^*_{00} = h_1(PO, ATC, DEP_0, RISK_0)$$
$$SR^*_{GG} = h_2(PG, ATC, DEP_G, RISK_G)$$
$$SIZE^*_{00} = z_1(PO, ATC, DEP_0, RISK_0)$$
$$SIZE^*_{GG} = z_2(PG, ATC, DEP_G, RISK_G)$$

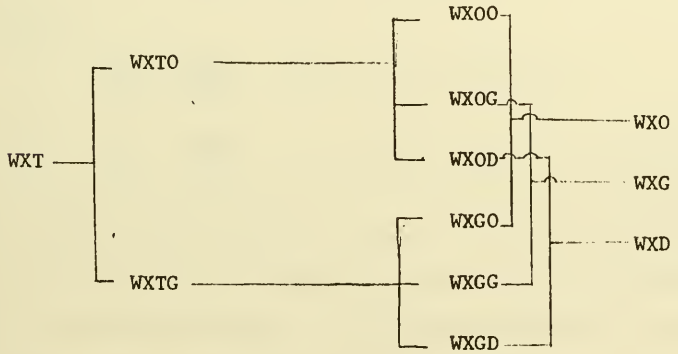
Issues in Aggregation

The discussion of the exploratory process thus far has centered around the number of wells intended for gas (or oil) and intended success ratios and sizes for gas (or oil) drilling. Unfortunately, it is extremely difficult, if not impossible, to obtain data on intentions of drillers. As a matter of fact, the only observable data are the results of the driller's decisions. Empirical data on wells is available only on the total number of exploratory wells drilled, numbers of successful gas wells and successful oil wells and the corresponding sizes of discovery per successful well under each category. We can therefore estimate only a single equation for total exploratory wells WXT rather than two separate equations for WXTO and WXTG. Similarly, we estimate equations for the ex-post relative fractions WXG/WXT and WXO/WXT rather than the ex-ante success ratios WXG/WXTG and WXO/WXTO.[1] Certain issues of aggregation become relevant in this context and are briefly touched upon below. Awareness of the specific nature of the aggregation problem will

[1]Note also that we can observe only the weighted averages of the results of the actions of all the explorers, i.e. the weighted average of the results of drilling all the prospects lying between the efficient inventory frontier and the critical profitability frontier. Thus in actuality, the position of the entire frontier in a given year and a given production district is represented by a single set of observations of total wells, success ratios, and sizes achieved.

prove to be useful in interpreting the results of the empirical estimation.

A preliminary categorization of the exploratory process on the basis of intention of drilling and results of drilling is given in the diagram below. In the diagram, "Characterization A", WXOO, WXOG, and WXOD represent oil wells intended for oil that turned out to be oil, gas and dry wells respectively. Similarly, WXGO, WXGG and WXGD represent categories of wells intended for gas. WXOG and WXGG together yield WXG, the observed total number of successful gas wells; WXOO and WXGO together make up WXO, the observed total number of successful oil wells.



"Characterization A"

If X is the vector of the relevant economic,geological and risk variables, we can write

$$\frac{WXTO}{WXT} = \Phi_O(X) \qquad \frac{WXTG}{WXT} = \Phi_G(X)$$

$$SR_{OO} = \frac{WXOO}{WXT} = h_1(X) \qquad SR_{GG} = \frac{WXGG}{WXTG} = h_2(X)$$

$$SR_{OG} = \frac{WXOG}{WXTO} = h_3(X) \qquad SR_{GO} = \frac{WXGO}{WXTG} = h_4(X)$$

$$\frac{WXOD}{WXTO} = 1-h_1(X)-h_3(X) \qquad \frac{WXGD}{WXTG} = 1-h_2(X)-h_4(X)$$

Note that $h_1(X)$ and $h_2(X)$ are the same as the "intended" success ratios SR^*_{OO} and SR^*_{GG} defined in the previous section. The functions $h_3(X)$ and $h_4(X)$ represent unintended (but probably anticipated) success ratios--$h_3(X)$ is the anticipated fraction of successful gas wells out of wells originally intended for oil, while $h_4(X)$ is the anticipated fraction of successful oil wells out of wells intended for gas.

We wish to estimate[1]

$$H_G(X) = \frac{WXG}{WXT} = \frac{WXGG}{WXTG} \cdot \frac{WXTG}{WXT} + \frac{WXOG}{WXTO} \cdot \frac{WXTO}{WXT}$$

$$= h_2(X) \cdot \Phi_O(X) + h_3(X) \cdot \Phi_G(X).$$

Similarly,

$$H_O(X) = \frac{WXO}{WXT} = h_1(X) \cdot \Phi_O(X) + h_4(X) \cdot \Phi_G(X)$$

and $$H_D(X) = \frac{WXD}{WXT} = [1-h_1(X)-h_2(X)]\ \Phi_O(X) + [1-h_3(X)-h_4(X)]\ \Phi_G(X)$$

What should be the directions of the effects of the predetermined variables on $H_G(X)$ and $H_O(X)$? Even assuming that expectations are realized, this question has an unambiguous answer only if certain conditions are satisfied. In the simplier case of perfect directionality, we have

$$H_O(X) = h_1(X) \cdot \Phi_O(X)$$

and $$H_G(X) = h_2(X) \cdot \Phi_G(X)$$

---

[1] In writing the above equations, we have implicitly assumed that on the whole and on the average, (i.e. when measured across many individual explorations and averaged through time), the expectations of the explorers are realized. This assumption would be a reasonable one when a large number of explorers are operating actively.

The effect of a marginal change of an arbitrary predetermined variable x is given by

$$\delta H_G(X)/\delta x = \Phi_G(X)\cdot(\delta h_2(X)/\delta x) + h_2(X)\cdot(\delta \Phi_G(X)/\delta x)$$

and $$\delta H_O(X)/\delta x = \Phi_O(X)\cdot(\delta h_1(X)/\delta x) + h_1(X)\cdot(\delta \Phi_O(X)/\delta x)$$

Here the levels of Φ_G, Φ_O, h_1 and h_2 are all positive and lie between 0 and 1. Thus if the signs of $\delta h_2(X)/\delta x$ and $\delta \Phi_G(X)/\delta x$ agree, there is no ambiguity in the sign of $\delta H_G(X)/\delta x$. The two depletion variables obviously would meet this condition since they have a negative effect on all the four functions Φ_G (X), $\Phi_O(X)$, $h_2(X)$ and $h_1(X)$. The effect of relative price PG/PO on Φ_G and Φ_O have predicatable signs in that a higher price ratio should increase Φ_G and decrease Φ_O. But the effects of prices on the intended success ratios $h_1(X)$ and $h_2(X)$ will be positive or negative depending on whether the extensive mode or intensive mode of drilling is predominant. In case the intensive mode dominates, $\delta h_1(X)/\delta PO>0$ and $\delta h_2(X)/\delta PG>0$, hence the signs of $\delta H_G/\delta PG$ and $\delta H_O/\delta PO$ are unambiguously positive. If the extensive mode dominates, $\delta h_1(X)/\delta PO$ and $\delta h_2(X)/\delta PG$ are negative, and so the signs of $\delta H_G(X)/\delta PG$ and $\delta H_O(X)/\delta PO$ could go either way. The sign of the ATC variable's influence also depends on which mode of drilling is more dominant. If the intensive mode is dominant, the cost should have a negative effect on the success ratios; the reverse would be true if drilling is predominantly extensive. We would however place less confidence on the estimated coefficients of this variable because no information is available on what portion of the cost is chargeable to gas exploration and what portion to oil exploration. High values of the risk variables should be associated with high values of the corresponding intended average sizes and success ratios. But high value of

risk will also have a negative effect on the corresponding relative fraction of wells intended -- for instance, higher values of $RISK_G$ cause a lesser proportion of wells to be intended for gas, thus leading to a higher $\Phi_G(X)$ and lower $\Phi_O(X)$. The net result of these two opposing influences can be determined only on the basis of empirical evidence.

In the absence of perfect directionality, the "unintended" success ratios $h_3(X)$ and $h_4(X)$ start to play a role in addition to the intended success ratios $h_1(X)$ and $h_2(X)$. Effects of economic variables on $h_3(X)$ and $h_4(X)$ are ambiguous in a magnitude as well as direction. However, in view of the evidence of high directionality and so the impact of $h_3(X)$ and $h_4(X)$ on $H_G(X)$ and $H_O(X)$ is not likely to be very large. There is no ambiguity in the expected signs of the depletion variables even in case of imperfect directionality because they affect all the functions in the same direction.

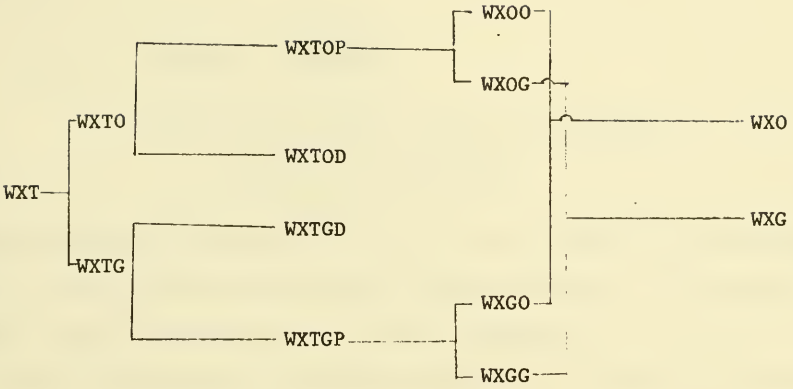
We are faced with similar problems of aggregation in estimating functional relationships for observed sizes of discoveries instead of intended sizes. Fortunately, the problem is much less severe in the case of size estimations because the influences of predetermined variables on the intended sizes are not confounded with their effects on the relative fraction of wells intended for gas and oil (Φ_G and Φ_O). Thus, at least under the assumption of perfect directionality, the observed sizes will be equal to the intended sizes. As a result of this, we can predict with greater confidence that higher values of $RISK_G$ and $RISK_O$ would be associated with higher average values of SIZEN and SIZEA respectively. As usual, the depletion variables should have a negative influence on the corresponding size variables. The own price effects on intended sizes (and hence observed sizes) would be negative while cost and cross-price effects would be positive if intensive drilling dominates. The reverse would be true if the predominant mode of drilling is extensive. For reasons already explained, we place more confidence in the predictions for the price coefficients than

those for the cost (ATC) coefficients. To the extent that there is less than perfect directionality, the signs could deviate from those mentioned above, but once again, we hope that the generally high observed directionalities will prevent this from being a major hurdle to our estimation.[1]

There is an alternative hypothesis regarding drilling decisions. According to this, it is suggested that even though drillers have a notion of whether they are looking for gas or oil, they merely look at the probability of finding either gas or oil. Under such an assumption, it is more appropriate to estimate the probabilities of finding either gas or oil in the first stage and then split up the successful wells into fractions of gas and oil wells in a second stage. Under this hypothesis, one should estimate an equation for (WXG + WXO)/WXT in the first stage and for WXG/(WXG + WXO) in the second stage. The process under this logic can be looked at better by adopting a slightly different characterization of wells, called "Characterization B" below.

---

[1] A related aggregation problem that is ignored here is that the motivation to drill may vary across different categories of wells--new pool and new field wildcats, deeper pool tests, shallower pool tests and extension wells--as a result of different risk characteristics. This disaggregated form will be handled in a later note.



"Characterization B"

In this case, we can once again write:

$$\frac{WXTOP}{WXTO} = g_1(X) \qquad WXOO = g_3(X)$$

$$\frac{WXTOD}{WXTO} = 1-g_1(X) \qquad \frac{WXOG}{WXTOP} = 1-g_3(X)$$

$$\frac{WXTGP}{WXYG} = g_2(X) \qquad \frac{WXGO}{WXTGP} = g_4(X)$$

$$\frac{WXTGD}{WXTG} = 1-g_2(X) \qquad \frac{WXGG}{WXTGP} = 1-g_4(X)$$

$$\frac{WXTO}{WXT} = \Phi_O(X) \qquad \frac{WXTGQ}{WXT} = \Phi_G(X)$$

We wish to estimate:

$$H_{GO}(X) = \frac{WXG+WXO}{WXT} = \frac{WXTOP+WXTGP}{WXT} = g_1(X)\cdot\Phi_O + g_2(X)\Phi_G$$

The fraction of gas wells in the total successful wells can then be estimated as the second stage of "Characterization B". This fraction is

given by

$$H_{G/GO}(X) = \frac{WXG}{WXG+WXO} = \frac{WXGG+WXOG}{WXTOP+WXTBP}$$

$$= \frac{g_1(X)\cdot(1-g_3(X))\Phi_O(X) + g_2(X)\cdot(1-g_4(X))\Phi_G(X)}{g_1(X)\Phi_O(X) + g_2(X)\Phi_G(X)}$$

The implications of aggregation in this case are similar to those discussed for the case of "Characterization A", but not identical. Once again, we can hope for sign preservation under certain conditions.[1]

Table 1 summarizes the predicted signs of the effects of the predetermined variables on the various functions under the assumption of perfect directionality. The direction of the corresponding effects on the fractions to be estimated H_G, H_O, H_{GO}, $H_{G/GO}$ are also given when ever they can be unambiguously predicted. Where the aggregation problems cause the signs to be ambiguous, a "?" is entered.

The issue of which of the two characterizations is a better description of reality can be settled only through empirical tests of alternative hypotheses. This is one of the issues we hope to deal with through econometric estimations.

---

[1] The sign is actually harder to predict for the second stage *viz*. $H_{G/GO}(X)$. For, the expression for WXG/WXGO can be written as:

$$\{\frac{(1-g_4(X))g_2(X)\Phi_G(X) + (1-g_3(X))g_1(X)\Phi_O(X)}{g_2(X)\Phi_G(X) + g_1(X)\Phi_O(X)}\}$$

$$= 1-g_4(X) - \frac{(g_3(X)-g_4(X)g_1(X)\Phi_O(X)}{g_2(X)\Phi_G(X) + g_1(X)\Phi_O(X)}$$

$$\frac{\delta\{\ \}}{\delta x} = (g_2(X)\Phi_G(X) + g_1(X)\Phi_O(X))(g_3(X) - g_4(X))g_1(X)(\delta\Phi_O/(X))$$

$$- (g_3(X) - g_4(X)g_1(X)\Phi_O(X)[g_2(X)\frac{\delta\Phi_G(X)}{\delta x} + g_1(X)\frac{\delta\Phi_O(X)}{\delta x}]/\ g_2(X)\Phi_G(X) + g_1(X)\Phi_O(X)$$

One advantage of the "double-binomial" logit formulation is the relative ease in interpreting the equation intuitively. In the "trinomial" formulation, the success ratio equations have to be used in conjunction with the constraint that the sum of the gas, oil and dry wells add up to total exploratory wells. Thus, without actually "simulating" the equations together it is difficult to interpret the effect of a change in one of the independent variables. Further, the equations used in both formulations consist of sets of non-linear equations and the choice as to which set of equations fits the actual situation is an empirical matter.

Derivation of the Functional Forms for the Success Ratio Estimations

The success ratios we wish to estimate are the relative fractions H_G, H_O and H_D (Characterization A) or H_{GO} and $H_{G/GO}$ (Characterization B). Note that all these fractions must lie in the zero-one interval by definition. However, each of the predetermined variables in these estimations is unconstrained; the function on the right hand side is unconstrained. There are well-known econometric biases and estimation difficulties involved in estimating such a relationship directly: to avoid this, we employ the so-called logit models as described below (15,17). We write the function on the right hand side as a composite function $g(f(X))$ where $f(X)$ is an unconstrained function of an attribute vector X (vector of predetermined variables) and g is a monotonic function that maps the range of $f(X)$ into a number between zero and one.[1]

[1] Without such a transformation, the equation is termed the linear probability model. It is essential for the validity of this specification that the linear function not range outside the zero one interval. For a discussion of this model see A.S. Goldberger, Econometric Theory (1964).

For instance, let $H_G = \frac{WXG}{WXT} = g(f(X))$. An obvious choice for the g-function is an ogive (cumulative distribution function) because it automatically satisfies the above mapping requirement. As it turns out in the present case, a cumulative distribution function is actually a natural choice for the g-function. To see this, one only has to observe the form of the $H_G(X)$ function derived from the theory of the individual explorer's choice behavior. This was given as:

$$H_G(X) = \frac{WXG}{WXT} = h_3(X)\Phi_0(X) + h_2(X)\Phi_G(X)$$

Here both $\Phi_0(X)$ and $\Phi_G(X)$ are cumulative distribution functions (cdf's) and satisfy $\Phi_0 + \Phi_G = 1$. Further $h_2(X)$ and $h_3(X)$ always lie between 0 and 1 and are relatively steady except under very large changes in the attribute vector X. We can therefore expect that $H_G(X)$ can be approximated by a cdf similar to $\Phi_0(X)$ and $\Phi_G(X)$.

While many alternative cdf's (such as cumulative normal, cumulative gamma, etc.) could have been used as our g-functions, those in the class of cumulative logistic distribution functions are chosen for reasons of convenience. The advantage of using these probability functions is that they yield computationally convenient forms that can be directly estimated by least squares or maximum likelihood methods. In the case of "Characterization B", the problem involves only binary choice and hence a simple logit probability model is appropriate. This uses the logistic distribution

$$H(X) = \frac{1}{1+e^{-f(X)}}$$

Using this,

$$1 - H(X) = \frac{e^{-f(X)}}{1+e^{-f(X)}}$$

which yields

$$\frac{H(X)}{1-H(X)} = e^{f(X)}$$

or $$\log \left(\frac{H(X)}{1-H(X)}\right) = f(X)$$

This is the logit form to be estimated. For least squares regression, we usually define $f(X)$ as a linear function βX.

In the case of "Characterization A", the decision tree has three prongs and hence a multinominal logit probability model is appropriate. Here, for the i^{th} alternative, define

$$H_i(X_i) = \frac{e^{f_i(X_i)}}{\sum_{j=1}^{J} e^{f_j(X_j)}}$$

where J is the total number of alternatives. In this case, it can be shown that

$$\log \left(\frac{H_i(X_i)}{H_j(X_j)}\right) = f_i(X_i) - f_j(X_j)).$$

In our present case, the attribute vector is the same for all the three alternatives and hence can be replaced by X, yielding

$$\log \left(\frac{H_i(X)}{H_j(X)}\right) = f_i(X) - f_j(X)$$

The equations appearing above can be directly estimated by least squares regression for any desired choice of the functions on the right hand side as

long as the chosen f(X) are linear on its parameters.[1] Assuming this, we can write:

$$\log\left(\frac{H_i(X)}{H_j(X)}\right) = (\beta_i - \beta_j)\, X$$

This is the form to be estimated under "Characterization A."

In addition to the attractive features of eliminating econometric biases and its natural flow from the explorer's choice problem, the logit functional forms offer some other advantages. It offers a very convenient way of estimating a non-linear functional relationship. For one thing, the forms directly embody the feature that the right hand side variables have a decreasing marginal effect on the dependent variable on either side of the median values. The so-called "S-curve effect", reflecting depletion of resources in a finite resource base, is incorporated in a natural way. Finally, the logit function is roughly linear around the middle of the range of the dependent variable so that the magnitudes and signs of the elasticities can be easily interpreted.

Thus the success ratio equations to be estimated under "Characterization A" are

---

[1] For more detailed exposition of the multinominal logit models, see McFadden (15) and Theil (17).

$$\log\left(\frac{H_G(X)}{H_O(X)}\right) = \log\left(\frac{WXG}{WXO}\right) = f_1(PG,\ PG/PO,\ ATC,\ DEP_G,\ DEP_O,\ RISK_G,\ RISK_O)$$

$$\log\left(\frac{H_O(X)}{H_D(X)}\right) = \log\left(\frac{WXO}{WXD}\right) = f_2(PG,\ PG/PO,\ ATC,\ DEP_O,\ RISK_O)$$

$$\log\left(\frac{H_G(X)}{H_D(X)}\right) = \log\left(\frac{WXG}{WXD}\right) = f_3(PG,\ PG/PO,\ ATC,\ DEP_G,\ RISK_G)$$

Actually, only two out of these three equations are independent and it suffices to estimate any two of the three. Using PG and PO together in the same equation is avoided to circumvent multi-collinarity problems as well as to facilitate separation of price effects on relative fractions of wells intended for gas or oil (Φ_G or Φ_O) from those on intended success ratios and sizes. The relative price term PG/PO is used in the above formulation to capture shifts in the relative fraction of wells intended for gas (or oil) while the own price terms indicate the effects on intended success ratios and sizes (thus giving a clue as to whether the extensive mode or intensive mode is predominant). Under "Characterization" B, the equations to be estimated are the two binomial logit forms.

$$\log\left(\frac{H_{GO}}{1-H_{GO}}\right) = \log\left(\frac{WXG+WXO}{WXD}\right) = f_4(PG,\ PG/PO,\ ATC,\ DEP_G,\ DEP_O,\ RISK_G,\ RISK_O)$$

and $$\log\left(\frac{H_{G/GO}}{1-H_{G/GO}}\right) = \log\left(\frac{WXG}{WXO}\right) = f_5(PG,\ PG/PO,\ ATC,\ DEP_G,\ DEP_O,\ RISK_G,\ RISK_O)$$

Functional Forms for the Average Size Estimations

The average size equations form the third segment of the model determining the levels of discoveries of non-associated and associated gas. Most of the non-associated reserve additions come from gas wells while associated reserve additions result primarily from oil wells. To take into account possible

differences in the characteristics of gas and oil wells, separate size equations are fitted for non-associated and associated reserves additions. This will facilitate capturing any differences in the way economic incentives influence the average size of associated and non-associated discoveries. Besides, fitting two different size equations allows us to verify structurally if a particular mode (extensive or intensive) of drilling dominates consistently. If either mode dominates consistently, we should find any decreases in the fraction of successes of gas wells in response to economic incentives to be accompanied by increases in the average size of non-associated discoveries and vice versa.

The relationships to be estimated for sizes of discoveries are:

Non-associated: $SIZEN = f_6(PG, PG/PO, ATC, DEP_G, RISK_G)$

Associated: $SIZEA = f_7(PO, PG/PO, ATC, DEP_O, RISK_O)$

Estimation of the Model

The model was estimated using pooled cross-section and time-series data. All the equations were estimated using data from 1964 to 1971. Data in earlier periods was available for many of the variables and could be approximated for the others but it was felt that the structural characteristics of the industry had changed considerably due to the imposition of regulation in 1961 (which took effect in the early sixties). All the equations were estimated over 20 Federal Power Commission districts.[1]

---

[1]The districts are California, Colorado-Utah, Kansas, Louisiana North, Louisiana South, Montana, Mississippi, New Mexico, Permian (including New Mexico South, Texas RRC districts 7C, 8 and 8A), Oklahoma, Pennsylvania, Texas RRC districts 1, 2, 3, 4, 6, 9, 10, West Virginia-Kentucky and Wyoming. Colorado and Utah are combined to form a single district. Similarly for West Virginia and Kentucky.

There are seven equations in the model - an equation for total exploratory wells drilled, two sets of success ratio equations, one each for the binomial and trinomial processes and a pair of average size equations for associated and non-associated gas respectively. The wells equation attempts to explain explanatory drilling for both oil and gas. The success ratios represent successful oil and gas wells as a ratio with respect to total exploratory drilling. The size variables SIZEN and SIZEA are the average sizes of discovery for non-associated and associated gas respectively.

In the case of the total wells equation, the log linear forms are preferred. In addition to the ease of interpretation of the elasticities they have the attractive feature that the implied price and cost elasticities are stable across regions and through time. This is resonable since exploration is carried on in many instances by the same firms operating in the various production districts. Thus drillers do not have rigid geographical preferences apart from the economic and geological factors already included in the model.

This explanation does not apply in the case of the success ratio equations, since the elasticities cannot be interpreted directly in the case of logit functional forms. The true estimate of the elasticities can be computed in these cases only by simulating the equations over the relevant period. Further, in these cases the use of a linear model is not inconsistent with roughly equal elasticities across regions and through time as the dependent variables do not change appreciably across regions. In view of these considerations both linear and log-linear

forms were tested on the data. This affords greater flexibility in capturing the non-linearities associated with various right hand side variables. When used in logit equations, both linear and log-linear forms incorporate the feature of decreasing marginal effects of the independent variables.

The economic variables used in the wells equation are REVD (deflated revenue per well),[1] ATCD (the deflated cost per exploratory well) and the interest rate INT. The (deflated) price of gas (PGD) and the price of oil (POD) are highly collinear during the period under study and cause difficulties of estimation. This problem is solved by using the REVD variable which is a composite variable made up of the two prices weighted by their relative importance, i.e. their levels of production in each district. Also, this enables us to separate the effects of substitution between oil and gas from the overall effect of economic incentives for exploration activity in general. The prices PGD and POD are used along with (PGD/POD) in the other equations in order to avoid the collinearity between the two price series. No geological variables are used in the wells equation since geographical variations caused by differences in the levels of drilling activity are already incorporated in the REVD variable. The geological differences and similarities across districts are taken into account in the size equations through the use of dummy variables DDA, DDB and DDC which account for heterogeneity among four broadly defined regional field markets.[2]

---

[1] All variables are deflated using the GNP price deflator.

[2] This was done on the basis of geological and market characteristics such as potential reserves and geographical proximity.

(The dummy variables are explained in Appendix A.) In the case of the success ratio equation it is felt desirable to use an independent variable that not only reflects the knowledge of the geological characteristics of the region but also provides a reference level for the dependent variable in that region. The elasticities implied by the other independent variables would thus be nearly equal across districts (which would not be true if we had used dummy variables). The geological variables chosen for the various equations are the historical means of the dependent variables. These are defined in detail in Appendix A.

The risk variables used in the different equations distinguish between districts in terms of the uncertainty faced by explorers and are constant through time in a particular district. The risk variable used in the wells equation is the variance of revenue per exploratory well (RISK). The higher this variance in a production district, the lower the incentive to drill exploratory wells in the district and hence we expect the sign of the RISK coefficient to be negative in the total wells equation.

The variances of the sizes of non-associated and associated gas discoveries per exploratory well $RISK_G$ and $RISK_O^1$ are the risk variables used in the success ratio and size equations. In accordance with the assumption of high directionality, the $RISK_G$ variables would be the appropriate index of risk while considering 'drilling for gas', while $RISK_O^1$ would be appropriate for 'drilling for oil.' Since oil wells yield associated gas as well as oil, a more appropriate risk variable for the equation predicting the fraction of oil wells would actually be the variance of the revenue from oil and associated gas per exploratory well drilled, $RISK_O$. An argument can be made that this is the more appropriate risk

variable in the equation for the size of associated discoveries. Both $RISK_0$ and $RISK_0^1$ are therefore tried as alternative possibilites in the equations where the size of associated discoveries or the fraction of oil wells appears as a dependent variable.

The interpretation of the risk terms in these equations is quite different from that in the wells equation. Based on the discussion of equilibrium in the market for oil and gas prospects, we expect higher average sizes SIZEN and SIZEA to be associated with higher values of the corresponding estimates of risks. $RISK_G$ and $RISK_0$(or $RISK_0^1$) and vice versa, given risk aversion by drillers. The explanation in the case of the success ratio equations is similar - we expect higher fractions of gas wells to be associated with higher values of $RISK_G$, while the fraction of oil wells is positively correlated with $RISK_0^1$. This discussion of the variables assumes that risks and returns accruing to oil and gas drilling are separable due to high directionality. The risk variables will add to the predictive power of the model only if the variances of the variables used tend to be more stationary than the respective means. This is because,as our discussion of the efficient frontier indicates, the mean and variance are to be treated simultaneaously rather than as a cause and effect.

The treatment of the depletion effects in these equations is somewhat different from past studies. For instance, Fisher takes account of the depletion of resources by including the lagged dependent variable as a right hand side variable. Based on his "inventory of prospects" hypothesis the coefficient of this variable should appear with a negative sign to indicate the effect of last year's drilling on the inventory of undrilled prospects. The Fisher (and Erickson-Spann) variable thus describes a short run phenomenon from one year to the next. The depletion variables used by us, on the con-

trary, attempt to capture a long run effect over several years due to the decline of the resource base itself.[1,2]

All the equations were estimated using weighted least squares regressions, to avoid problems of heteroscedasticity. The major problem in this context arises in the success ratio equations where the least squares assumption of homoscedasticity is isolated by the nature of the multinomial choice process. The process has the charateristic that the variance of the error term depends on the level of the relative frequencies of the various events. Expressions for the variances of the error terms as well as the appropriate weighting factors to be used in the success ratio equations are derived in Appendix B.

In the case of the total exploratory wells, the variance of the disturbance term is a function of the "size" of the district, i.e. the greater the number of wells drilled in a particular district, the larger the variance of the error. It is assumed that the variance of the disturbance is approximately proportional to the mean "size" of the district. The approximate weights used to transform the variables in order to satisfy the homoscedasticity assumptions are thus given by $1/\sqrt{\text{"size"}}$ where the indicator of "size" for our purposes is the mean number of wells drilled in the particular district through time. Similarly for the SIZEN and SIZEA equations,

---

[1]Following Fisher - E-S, we attempted to capture a one-period "inventory of prospects" depletion, but could not find any appreciable statistically significant effect.

[2]This also spares us the problems faced by Fisher in confounding the depletion effects with the district distinguishing effects, since the lagged variable in effect serves as a district factor. Furthermore, use of the lagged variable could introduce certain econometric biases into the OLSQ estimates.

the average reserve additions per well through time were taken as indicators of "size" and the appropriate weighting factors computed.

Statistical Results

The results of the estimation are presented in equations 1 through 7. The equations were estimated by weighted least squares. The numbers in parenthesis are the t-statistics of the respective coefficients. It should be pointed out that the Durbin-Watson statistic has limited meaning since the effects of auto-correlation through time and across cross-sections are confounded.

$$\begin{aligned} LOG(WXT) = {} & -1.437 + 0.829\ REVD(-1) + 0.404\ \frac{PGD(-1)}{POD(-1)} \\ & (-1.396)\ (17.89)\ (1.222) \\ & -0.214\ ATCD(-1)\ -0\ 833\ DEP_G(-1) \\ & (-2.317)\ (-2.627) \\ & -1.335\ DEP_O(-1)\ -0.252\ INT(-1) \\ & (-2.67)\ (-1.097) \\ & -0.168\ RISK \\ & (-6.904) \end{aligned} \qquad (1)$$

F ratio = 62.49 R^2 = 0.80 d.f. = 106

As expected, the (deflated) revenue per well enters with a strong positive sign. The capital costs represented by the interest rate and the deflated drilling cost variable ATCD act as negative economic indicators and have the expected negative impact on well drilling activity. The price ratio PGD/POD represents the substitution effect between total wells drilled for gas and those for oil. Ideally, we would have liked to estimate separate

equations for wells intended for gas and oil but such data is unavailable. The term (PGD/POD) is an attempt to separate the two types of wells by allowing the well drilling activity to be more sensitive to one price than the other. The positive sign of this coefficient seems to indicate that drilling activity in recent years has been more sensitive to shifts in gas prices rather than oil prices. But one cannot emphasise this overly as the coefficient is barely significant at the 10% level. Both the oil and gas depletion variables have statistically significant negative coefficients indicating that exploration activity adjusts for the depletion of resources. The coefficient of the risk variable has a negative sign indicating that other things being equal, well-drilling activity is lower in districts with greater uncertainty of revenues per well.

The Success Ratio Equations

Two sets of success ratio equations corresponding to "characterizations" A and B are presented below. Equations 2 and 3 are the estimations for Characterization A and flow from the trinomial logit model, while equations 4 and 5 are the corresponding equations for Characterization B and are formulated as two binomial logit equations.

Characterization A

$$\text{LOG}\left(\frac{WXG + 1/2}{WXD + 1/2}\right) = \underset{(3.289)}{45.671} - \underset{(-1.627)}{0.139}\ PGD(-1) - \underset{(-1.481)}{0.294}\ \frac{PGD(-1)}{POD(-1)}$$

$$\underset{(2.140)}{-3.78X10^{-6}}\ ATCD(-1) - \underset{(-4.504)}{2.05}\ DEP_G(-1) \qquad (2)$$

$$\underset{(4.653)}{+14.61}\ (GEOL_1) - \underset{(-0.700)}{1.203X10^{-7}}\ RISK_G$$

F ratio = 444.49 $\qquad R^2 = 0.96$ $\qquad$ d.f. = 107

$$\text{LOG}\left(\frac{WXO + 1/2}{WXD + 1/2}\right) = 20.8824 - 0.65113\ POD(-1) - 0.16283\ \frac{PGD(-1)}{POD(-1)} \quad (3)$$

(3.58192) (-3.65999) (-3.55873)

$$+2.4101X10^{-6}\ ATCD(-1) - 2.43671\ DEP_O(-1) + 7.59887(GEOL_2)$$

(-3.45174) (-2.91617) (7.05456)

$$-1.04139X10^{-8}\ RISK_O$$

(-1.08796)

F ratio = 136.072 R^2 = 0.88 d.f. = 107

Characterization B

$$\text{LOG}\left(\frac{WXGO + 1/2}{WXD + 1/2}\right) = -2.332 - 0.051\ PGD(-1) - 0.076\ \left(\frac{PGD(-1)}{POD(-1)}\right)$$

(-5.29) (-1.911) (1.035)

$$-2.02X10^{-6}\ ATCD(-1) - 1.394\ DEP_G(-1) \quad (4)$$

(-3.318) (-3.895)

$$-1.621\ DEP_O(-1) + 4.239(GEOL_3)$$

(-3.263) (9.202)

$$+6.369X10^{-5}\ RISK_G + 1.237\ RISK_O$$

(3.326) (3.167)

F ratio = 70.64 R^2 = 0.82 d.f. = 121

$$\text{LOG}\left(\frac{WXG + 1/2}{WXO + 1/2}\right) = 0.956 + 0.403\ PGD(-1) + 1.33\ \frac{PGD(-1)}{POD(-1)}$$

(0.678) (-0433) (1.708)

$$-0.167\ ATCD(-1) + 0.114\ DEP_G(-1)$$

(-1.167) (0.235)

$$+0.954\ DEP_O(-1) + 0.897(GEOL_4) \quad (5)$$

(1.461) (6.878)

$$+0.0307\ RISK_G + 0.0037\ RISK_O$$

(0.629) (0.06955)

F ratio = 37.90 R^2 = 0.74 d.f. = 105

The negative signs on the coefficients of the own price terms indicate that the "extensive" mode of drilling is more dominant, i.e. as price of gas (oil) increases, explorers shift towards less prospects that are less certain but with larger expected sizes of discoveries, conditional on success. The signs of the price ratio term (PGD/POD) reflect shifts in the relative fraction of exploratory wells originally intended for oil or gas. As PGD/POD increases, the fraction of total exploratory drilling intended for gas goes up. This causes the success ratio for gas wells to increase and the success ratio for oil wells to decline (assuming, of course, that directionality is high, i.e. a higher fraction of successful gas (oil) wells result from exploratory drilling for gas (oil) than from oil (gas)).

For the trinomial forms we observe a depletion effect for oil and gas, although not very significant (barely significant at the 10% level). The equilibrium hypothesis requires that the coefficients of the risk variables to be positive. This is not true in one of the equations but the coefficient is statistically insignificant.

In line with the evidence in favor of the dominance of extensive mode of drilling, we would expect to find positive signs on the coefficients of the average cost term in both the equations. The cost terms in one of the equations thus does not have the predicted sign. This may have been due to several reasons. Firstly, the cost date used included only the drilling costs whereas the appropriate cost to be used would be the total exploratory costs per well. Even the drilling cost date is reported only for aggregates of exploratory and developmental drilling. Inclusion of developmental drilling costs could substantially contaminate the cost figures. Finally, the data does not give us any information as to how much of the costs are

attributable to drilling for gas and how much for drilling for oil. The relative values of the costs attributable to gas and oil can shift without altering the estimate of the ATC. It is probably best to regard the ATC as a variable whose primary function is to filter the effects of district to district variations in costs (thus leading to better estimates of the other variables' effects) rather than give it a strong economic interpretation.

The first equation of the double binomial formulation has both the depletion variables as well as both the risk variables as independent variables since the left hand side refers to both oil and gas wells. In the first equation of the double binomial formulation, both the depletion variables come out strongly with the expected negative signs while the risk variables show the expected positive signs. Note that in the second equation of the double binomial formulation, the expected sign for DEP_0 would be positive because the number of oil wells appears in the denominator of the left hand side. For the same reason, the expected sign on the risk variable $RISK_0$ would be negative for the second equation in the double binomial formulation. We thus see that the signs of the risk variables in this equation are not as predicted by our theory, but both the coefficients are statistically insignificant.

The Size Equations

The estimated equations for the average sizes of discovery, both associated and non-associated are presented overleaf.

The significantly positive own price effects on the sizes of both associated and non-associated discoveries lend further support to the claim that the extensive mode of drilling is the predominant one. The explorers react to a price increase by shifting to less certain but larger sized prospects. We see further evidence of depletion in both oil and gas reserves in these equations.

$$
\begin{aligned}
LOG(SIZEN) = {} & -0.253\ DDA + 0.628\ DDB - 0.406\ DDC - 1.379 \\
& (-0.416) \quad (1.391) \quad (-1.885) \quad (-0.360) \\
& +2.521\ LOG(PGD(-1)) - 2.580\ LOG(PGD(-1)/POD(-1)) \\
& (1.547) \quad (-0.654) \\
& +0.340\ LOG(ATCD(-1)) - 0.632\ LOG(DEP_G(-1)) \\
& (1.279) \quad (-1.221) \\
& +0.376\ LOG(RISK_G) \\
& (6.127)
\end{aligned}
\tag{6}
$$

F Ratio = 11.4 $\quad R^2 = 0.47 \quad$ d.f. = 105

$$
\begin{aligned}
LOG(SIZEA) = {} & -0.092\ DDA - 1.454\ DDB + 0.170\ DDC \\
& (-0.112) \quad (-2.023) \quad (0.576) \\
& -11.874 + 2.519\ LOG(POD(-1)) - 2.697\ LOG(PGD(-1)/POD(-1)) \\
& (-2.689) \quad (1.424) \quad (-2.858) \\
& +1.343\ LOG(ATCD(-1)) - 0.165\ LOG(DEP_O) \\
& (4.010) \quad (-0.130) \\
& +0.351\ LOG(RISK_O) \\
& (3.434)
\end{aligned}
\tag{7}
$$

F Ratio = 9.101 $\quad R^2 = 0.44 \quad$ d.f. = 91

As in the case of the success ratio equations, the price ratio (PGD/POD) was expected to have a positive effect on the size of non-associated gas discoveries and a negative effect on the size of associated gas discoveries. We observe the expected signs in the equation for associated gas discovery size but not in the corresponding non-associated equation. Once again, the ATCD term cannot be given a strict economic interpretation.

		of intended gas/oil wells		Intended success ratios		Success fractions				(and observed) average size	
		$\phi_G(X)$	$\phi_O(X)$	$h_1(X)$	$h_2(X)$	$H_G(X)$	$H_O(X)$	$H_{GO}(X)$	$H_{G/GO}(X)$	$Z_1(X)$ (SIZEN)	$Z_2(X)$ (SIZEA)
PG	Intensive mode	0	0	+	0	+	0	+	+	-	0
	Extensive mode	0	0	-	0	-	0	-	-	+	0
PO	Intensive mode	0	0	0	+	0	+	+	-	0	-
	Extensive mode	0	0	0	-	0	-	-	+	0	+
PG/PO	Intensive mode	+	-	0	0	+	-	?	+	?	?
	Extensive mode	+	-	0	0	+	-	?	+	?	?
ATC	Intensive mode	?	?	-	-	-	-	-	?	+	+
	Extensive mode	?	?	+	+	+	+	+	?	-	-
DEP_G		-	+	-	0	-	0	-	-	-	0
DEP_O		+	-	0	-	0	-	-	+	0	-
$RISK_G$		-	+	+	-	?	?	+	?	+	0
$RISK_O$		+	-	-	+	?	?	+	?	0	+

Table 1

Summary of predicted signs of the effects of the right hand side variables for the case of perfect directionality

Note: A '+' indicates an expected positive effect, a '-' indicates an expected negative effect and a '0' is inserted when no significant effect in either direction is predicted. In cases where a significant effect may be present, but its direction is ambiguous because of aggregation problems, a '?' is inserted.

References

1. K.J. Arrow, Essays on the Theory of Risk Bearing, Markham Publishing Co., Chicago (1971).

2. P. Balestra and M. Nerlove, "Pooling Cross-Section and Rime Series Data in the Estimation of a Dynamic Model: the Demand for Natural Gas, Econometrica, July 1966.

3. E. W. Erickson and R. M. Spann, "Supply Price in a Regulated Industry, the Case of Natural Gas," Bell Journal of Economics and Management Science, Vol. 2, No. 1 (1971).

4. E. W. Erickson and R. M. Spann, "Price, Regulation and the Supply of Natural Gas in the United States" in Regulation of the Natural Gas Producing Industry, Keith C. Brown, ed., Resources for the Future, Incorporated, 1970.

5. F. M. Fisher, Supply and Costs in the U.S. Petroleum Industry: Two Econometric Studies, Johns Hopkins Press (Baltimore, 1964)

6. T. M. Garland and W. D. Dietzman, "Engineering Cost Study of Development Wells and Profitability Analysis of Crude Oil Production," U.S. Bureau of Mines Information Circular #8561, 1972.

7. G. Kaufman "Sampling Without Replacement and Proportional to Random Size," Memorandum II, to the MIT Natural Gas Research Group, March 19, 1973.

8. J. D. Khazzoom, "Gas Production Directionality", Public Utilities Fortnightly, Vo. 84 (December 18, 1969).

9. J. D. Khazzoom, An Econometric Model of U.S. Natural Gas Supply, Testimony before the Federal Power Commission in FPC Docket AR69-1, FPC Area Rate Proceedings.

10. J. D. Khazzoom, "The F.P.C. Staff's Econometric Model of Natural Gas Supply in the United States," Bell Journal of Economics and Management Science, Vol. 2, No. 1 (1971).

11. F. H. Lahee, Statistics of Expiratory Drilling in the United States, American Association of Petroleum Geologists, 1962.

12. P. W. MacAvoy, Price Information in Natural Gas Fields, Yale University Press (New Haven, 1962).

13. P. W. MacAvoy, "The Regulation-Induced Shortage of Natural Gas," Journal of Law and Economics, April 1971.

14. P. W. MacAvoy and R. S. Pindyck, "Alternative Regulatory Policies for Dealing with the Natural Gas Shortage," Bell Journal of Economics and Management Science, Vol 4, No. 2, Fall, 1973.

15. D. McFadden, "Conditional Logit Analysis of Qualitative Choice Behavior," Working Paper No. 199/BART 10, University of California, Berkeley, (November, 1972).

16. P. A. Samuelson, "Fundamental Approximation Theorem of Portfolio Analysis in terms of Means, Variances and Higher Order Moments," Review of Economic Studies, October 1970.

17. H. Theil, "A Multinominal extension of Linear Logit Model," International Economic Review, 1969.

18. J. Tobin, "Liquidity Preference as Behavior Towards Risk," Review of Economic Studies, February, 1958.

19. R. M. Spann and E. W. Erickson, "Joint Costs and Separability in Oil and Gas Exploration" in Energy Modelling, Milton Searl, Ed., Resources for the Future, Inc., 1973.

20. J. Tobin, "Liquidity Preference as Behavior Towards Risk," Review of Economic Studies, February, 1958.

APPENDIX A

List of Variables and Data Sources

The variables used to estimate the model, including sources of data, are listed below. The subscripts j, t indicating the index for the j^{th} production district for the t^{th} period have been omitted unless the definition requires such mention.

Reserves: All data from American Gas Association/American Petroleum Institute/Canadian Petroleum Association/Reserves of Crude Oil in disaggregated (i.e. associated-dissolved, non-associated, by production district) form from 1966 except for a year-end reserves (YN,YA,US) which we have from 1965. Data is given in millions of cubic feet.

SIZEN: Average size of non-associated discoveries per successful gas well by production district.

SIZEA: Average size of associated discoveries per successful oil well by production district.

DDA: Dummy variable for Louisiana south.
DDB: Dummy variable for Permian producing region.
DDC: Dummy variable for Kansas, Oklahoma, TRRC Districts 1,2,3,4, and 10.
DDD: Dummy variable for Louisiana north, Mississippi, New Mexico north, Pennsylvania, West Virginia-Kentucky, Wyoming, TRRC6 and TRRC9.
YN: Year end non-associated reserves. Reserves as defined by the AGA.
YA: Year end associated-dissolved reserves. See YN.
YT: Year end total gas reserves. YT = YN + YA.
YO: Year end oil reserves.
XN: Extensions of non-associated gas. Includes any newly proved reserves already established in pools in fields.
XA: Extensions of associated-dissolved gas. See XN.
DN: Total new discoveries, non-associated.
DA: Total new discoveries, associated-dissolved gas.

Wells:

WXT = Total number of exploratory wells, from the Joint Association Survey of Drilling Statistics.
WXO = Number of successful oil wells.
WXG = Number of successful gas wells.
WXD = Number of dry wells.

Success Ratios:

SR_G = Success ratio for gas wells = WXG/WXT.

SR_O = Success ratio for oil wells = WXO/WXT.

Prices and Economic Variables:

PG: New contract price of interstate sales of gas at the wellhead, for 1952-1971, cf. (Compiled by Foster Associates, Inc. Washington, D.C.)

PO: Wellhead price of oil in $/barrel for 1954-1971, from Bureau of Mines, Minerals Yearbook.

REV: Index of average revenue from gas and oil for each production district.

ATC: Average drilling cost per well of exploratory and development wells, from Joint Association Survey (AGA/API/CPA).

PGD, POD, REVD and ATCD
Deflated variables (deflated by the GNP deflator) corresponding to the above.

Depletion Variables:

PGC_j = Potential Gas Committee's estimates of the total probable, possible and speculative reserves in production district j.

$$(DEP_G)_{t,j} = \frac{\text{[Estimated Original Gas-in-Place - Cumulative Gas Reserves recovered till time t] in district j.}}{\text{Estimated Original Gas-in-Place in district j.}}$$

$$(DEP_O)_{t,j} = \frac{\text{[Estimated Original Oil-in-Place - Cumulative Oil Reserves recovered till time t] in district j}}{\text{Estimated Original Oil-in-Place in district j}}$$

Risk Variables

RISK Index of the Variance of Dollar Returns from both Gas (Associated and Non-Associated) and Oil Discoveries per Exploratory Well Drilled, by production district.

$RISK_G$ Index of the variance of total Non-associated Gas Discoveries made for successful gas well, by production district.

$RISK_O$ Index of the variance of total revenues from discoveries of Associated Gas and Oil per successful oil well, by production district.

$RISK'_O$ Index of the variance of total Associated Gas Discoveries made per successful oil well, by production district.

Geological Variables:

$$(GEOL_1)_j = \frac{\text{Sum of all successful gas wells drilled since 1964 in district 'j'}}{\text{Sum of all dry wells drilled since 1964 in district 'j'}}$$

$$(GEOL_2)_j = \frac{\text{Sum of all successful oil wells drilled since 1964 in district 'j'}}{\text{Sum of all dry wells drilled since 1964 in district 'j'}}$$

$$(GEOL_3)_j = (GEOL_1)_j + (GEOL_2)_j$$

$$(GEOL_4)_j = \frac{(GEOL_1)_j}{(GEOL_2)_j}$$

APPENDIX B

Weights to Correct for Heteroscedasticity in the Logit Model

Binomial Model

We will consider a general binomial logit model which may be expressed as

$$\log \frac{p}{1-p} = \alpha + \beta X \qquad \text{A1}$$

where p is the true probability of the occurrence of a particular event and X is the set of independent variables that determine the level of p.

Let f be the observed relative frequency of this event as computed from n observations. The relative frequency f will not, in general, equal p due to sampling errors. Thus,

$$\log \frac{f}{1-f} = \alpha + \beta X + \varepsilon \qquad \text{A2}$$

where $\varepsilon = \log \left(\frac{f}{1-f}\right) - \log \left(\frac{p}{1-p}\right)$

We shall term ε the "observational error."

If we have time series-cross sectional observations of the relative frequencies, we have for the k^{th} production district and the t^{th} period,

$$\log \frac{f_{kt}}{1-f_{kt}} = \alpha + \beta X_{kt} + \varepsilon_{kt} \qquad \text{A3}$$

Since we wish to estimate A3 using ordinary least squares, the error terms ε_{kt} must satisfy the following three conditions:

(i) $E(\varepsilon_{kt}) = 0$

(ii) The error terms ε_{kt} must all have the same variance

(iii) The error terms must be pairwise uncorrelated.

If n_{kt} is the sample size on the basis of which f_{kt} is computed and if the process generating the event under consideration is assumed binomial,

for reasonably large n_{kt} the error term ε_{kt} will have zero expectation. If the populations from which the different samples are drawn are independ t, it is reasonable to assume that condition (iii) is satisfied. However, condition (ii) is not satisfied, because

$$\text{Var}\,[\varepsilon_{kt}] = E[\log(\frac{f_{kt}}{1-f_{kt}}) - \log(\frac{p_{kt}}{1-p_{kt}})]^2$$

$$\text{since } E[\varepsilon_{kt}] = 0 \ .$$

For sufficiently large n_{kt}, the distribution of f_{kt} is concentrated around p_{kt} and a linear Taylor expansion suffices:

$$\log(\frac{f_{kt}}{1-f_{kt}}) - \log(\frac{p_{kt}}{1-p_{kt}}) = (f_{kt} - p_{kt}) \cdot \frac{\partial L_{kt}}{\partial p_{kt}} \qquad \text{A4}$$

where L_{kt} is the logit function

$$L_{kt} = \log(\frac{p_{kt}}{1-p_{kt}})$$

$$\frac{\partial L_{kt}}{\partial p_{kt}} = \frac{\partial}{\partial p_{kt}} [\log p_{kt} - \log(1-p_{kt})] = \frac{1}{p_{kt}(1-p_{kt})}$$

From A4, $\quad \text{Var}\,(\varepsilon_{kt}) = \dfrac{1}{p_{kt}^2(1-p_{kt})^2} \cdot \text{Var}\,(f_{kt})$

For a binomial process,

$$\text{Var}\,(f_{kt}) = \frac{1}{n_{kt}} p_{kt}(1-p_{kt}) \ .$$

Hence,

$$\text{Var}\,(p_{kt}) = \frac{1}{n_{kt} \cdot p_{kt}(1-p_{kt})} \ .$$

Thus, for large n_{jt}, to achieve constant variances for the error terms, the $(k,t)^{th}$ observation must be assigned a multiplicative weight of

$$w_{kt} = \sqrt{n_{kt} \cdot f_{kt}(1-f_{kt})}$$

Multinomial Model

Consider the case of the multinomial logti function:

$$\log\left(\frac{p_i}{p_j}\right) = \alpha_{ij} + \beta_{ij}X$$

where p_i and p_j are the probabilities of occurrence of the i^{th} and j^{th} events and X is the set of independent variables that determine their relative probabilities. f_i and f_j are the observed relative frequencies of the i^{th} and j^{th} events as computed from n observations. The relative frequencies f_i and f_j differ from p_i and p_j due to sampling errors ε_{ij} as follows:

$$\log \frac{f_i}{f_j} = \alpha_{ij} + \beta_{ij}X + \varepsilon_{ij}$$

where $$\varepsilon_{ij} = \log\left(\frac{f_i}{f_j}\right) - \log\left(\frac{p_i}{p_j}\right)$$

If we have time series-cross sectional observations, we have for the k^{th} production district and the t^{th} period

$$\log \frac{f_{ikt}}{f_{jkt}} = \alpha_{ij} + \beta_{ij}X_{kt} + \varepsilon_{ijkt}$$

If n_{kt} is the sample size on the basis of which f_{ikt} and f_{jkt} are computed and if the process generating the event is assumed multinomial, the error term ε_{ijkt} will satisfy the leasts squares assumptions for reasonably large n_{kt}, except for the condition that the error terms have the same variance.[@]

$$\text{Var}\ (\varepsilon_{ij}) = E\left[\log\left(\frac{f_i}{f_j}\right) - \log\left(\frac{p_i}{p_j}\right)\right]^2$$

Let $$L^{ij} = \log\left(\frac{p_i}{p_j}\right)$$

Taking the Taylor expansion of $\log\left(\frac{f_i}{f_j}\right) - \log\left(\frac{p_i}{p_j}\right)$

[@]The subscripts kt are dropped for simplicity of notation.

we have $$\log\left(\frac{f_i}{f_j}\right) - \log\left(\frac{p_i}{p_j}\right) = (f_i - p_i)\frac{\partial L^{ij}}{\partial p_i} - (f_j - p_j)\frac{\partial L^{ij}}{\partial p_j}$$

$$= (f_i - p_i)\frac{1}{p_j} - (f_j - p_j)\frac{p_i}{p_j^2}$$

Hence $$\mathrm{Var}(\varepsilon_{ij}) = E\left[\left(\frac{f_i - p_i}{p_j}\right) - \left(\frac{f_j - p_j}{p_j}\right)\left(\frac{p_i}{p_j}\right)\right]^2$$

$$= E\left[\left(\frac{f_j - p_i}{p_j}\right) + \left(\frac{f_j - p_j}{p_j}\right)^2 \frac{p_i^2}{p_j^2} - 2\frac{(f_i - p_i)(f_j - p_j)}{p_j^2}\frac{p_i}{p_j}\right]$$

$$= \frac{1}{p_j^2}\mathrm{Var}(f_i) + \frac{1}{p_j^2}\left(\frac{p_i^2}{p_j^2}\right)\mathrm{Var}(f_j) - 2\cdot\frac{1}{p_j^2}\left(\frac{p_i}{p_j}\right)\mathrm{Cov}(f_i, f_j)$$

$$= \frac{1}{np_j^2}\left[\frac{p_i(p_i + p_j)}{p_j}\right] = \frac{p_i^2}{np_j^2}\frac{(p_i + p_j)}{(p_i\, p_j)}$$

The appropriate weights to satisfy the least squares assumptions are given by

$$w_{ij} = \sqrt{\frac{1}{\mathrm{Var}(G_{ij})}}$$

$$= \frac{\sqrt{n}\sqrt{p_i p_j}}{\sqrt{p_i + p_j}}\left(\frac{p_j}{p_i}\right)$$

which, for large n, can be written as

$$\frac{\sqrt{n}\sqrt{f_i f_j}}{\sqrt{f_i + f_j}}\left(\frac{f_j}{f_i}\right)$$

Date Due

SEP 21 '78

Lib-26-67

MIT LIBRARIES
3 9080 003 702 591
716-74

MIT LIBRARIES
3 9080 003 702 617
717-74

MIT LIBRARIES
3 9080 003 702 633
718-74

MIT LIBRARIES
3 9080 003 702 658
719-74

HD28.M414 no.720- 74
Bloom, Gordon /Corporate social respon
800385 D*BKS 00037714
3 9080 000 866 738

MIT LIBRARIES
3 9080 003 702 674
721-74

HD28.M414 no.722- 74
Madnick, Stuar/Application and analysi
721943 D*BKS 00028908
3 9080 000 761 764

BOUND BY HAM